Estimation in autoregressive model with measurement error

Jérôme Dedecker, Adeline Samson, Marie-Luce Taupin

ESTIMATION IN AUTOREGRESSIVE MODEL WITH MEASUREMENT ERROR

JÉRÔME DEDECKER[(1)], ADELINE SAMSON[(1)], MARIE-LUCE TAUPIN[(2)]

(1) Laboratoire MAP5 UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité,
(2) Laboratoire Statistique et Génome UMR CNRS 8071- USC INRA, Université d'Évry Val d'Essonne

Abstract. Consider an autoregressive model with measurement error: we observe $Z_i = X_i + \varepsilon_i$, where X_i is a stationary solution of the autoregressive equation $X_i = f_{\theta^0}(X_{i-1}) + \xi_i$. The regression function f_{θ^0} is known up to a finite dimensional parameter θ^0. The distributions of X_0 and ξ_1 are unknown whereas the distribution of ε_0 is completely known. We want to estimate the parameter θ^0 by using the observations $Z_0, \ldots, Z_n$. We propose an estimation procedure based on a modified least square criterion. This procedure provides an asymptotically normal estimator $\hat{\theta}$ of θ^0, for a large class of regression functions and various noise distributions.

Keywords: autoregressive model, Markov chain, mixing, deconvolution, semi-parametric model.
AMS 2000 MSC: Primary 62J02, 62F12, Secondary 62G05, 62G20.

1. Introduction

We consider an autoregressive model with measurement error satisfying

$$\left\{ \begin{array}{ll} Z_i & = X_i + \varepsilon_i, \\ X_i & = f_{\theta^0}(X_{i-1}) + \xi_i \end{array} \right. \tag{1.1}$$

where one observes $Z_0, \cdots, Z_n$ and the random variables $\xi_i, X_i, \varepsilon_i$ are unobserved. The regression function f_{θ^0} is known up to a finite dimensional parameter θ^0, belonging to the interior Θ° of a compact set $\Theta \subset \mathbb{R}^d$. The centered innovations $(\xi_i)_{i\geq 1}$ and the errors $(\varepsilon_i)_{i\geq 0}$ are independent and identically distributed (i.i.d.) random variables with finite variances $\mathrm{Var}(\xi_1) = \sigma_\xi^2$ and $\mathrm{Var}(\varepsilon_0) = \sigma_\varepsilon^2$. We assume that ε_0 admits a known density with respect to the Lebesgue measure, denoted by f_ε. Furthermore we assume that the random variables X_0, $(\xi_i)_{i\geq 1}$ and $(\varepsilon_i)_{i\geq 0}$ are independent. The distribution of ξ_1 is unknown and does not necessarily admit a density with respect to the Lebesgue measure. We assume that $(X_i)_{i\geq 0}$ is strictly stationary, which means that the initial distribution of X_0 is an invariant distribution for the transition kernel of the homogeneous Markov chain $(X_i)_{i\geq 0}$.

Our aim is to estimate θ^0 for a large class of functions f_θ, whatever the known error distribution, and without the knowledge of the ξ_i's distribution. The distribution of the innovations being unknown, this model belongs to the family of semi-parametric models.

Previously known results. Several authors have considered the case where the function f_θ is linear (in both θ and x), see *e.g.* Andersen and Deistler (1984), Nowak (1985), Chanda (1995, 1996), Staudenmayer and Buonaccorsi (2005), and Costa *et al.* (2010). We can note that, in this specific case, the model (1.1) is also an ARMA model (see Section 4.1.1 for further details). Consequently, all previously known estimation procedures for ARMA models can be applied here, without assuming that the error distribution is known.

For a general regression function, the model (1.1) is a Hidden Markov Model with possibly a non compact continuous state space, and with unknown innovation distribution. When the innovation distribution is known up to a finite dimensional parameter, the model (1.1) is fully parametric and various results are already stated. Among others, the parameters can be estimated by maximum likelihood, and consistency, asymptotic normality and efficiency have been proved. For further references on estimation in fully parametric Hidden Markov Models, we refer for instance to Leroux (1992), Bickel *et al.* (1998), Jensen and Petersen (1999), Douc and Matias (2001), Douc *et al.* (2004), Fuh (2006), Genon-Catalot and Laredo (2006), Na *et. al.* (2006), and Douc *et al.* (2011).

In this paper, we consider the case where the innovation distribution is unknown, and thus the model is not fully parametric. In this general context, there are few results. To our knowledge, the only paper which gives a consistent estimator is the paper by Comte and Taupin (2001). These authors propose an estimation procedure based on a modified least squares minimization. They give an upper bound for the rate of convergence of their estimator, that depends on the smoothness of the regression function and on the smoothness of f_ε. Those results are obtained by assuming that the distribution P_X of X_0 admits a density f_X with respect to the Lebesgue measure and that the stationary Markov chain $(X_i)_{i\geq 0}$ is absolutely regular (β-mixing). The main drawback of their approach is that their estimation criterion is not explicit, hence the links between the convergence rate of their estimator and the smoothness of the regression function and of the error distribution are not explicit either. Consequently, Comte and Taupin (2001) are able to prove that their estimator achieves the parametric rate only for very few couples of regression functions/error distribution. Lastly their dependency conditions are quite restrictive, and the assumption that X admits a density is not natural in this context.

Our results. In this paper, we propose a new estimation procedure which provides a consistent estimator with a parametric rate of convergence in a very general context. Our approach is based on the new contrast function

$$S_{\theta^0,P_X}(\theta) = \mathbb{E}[(Z_1 - f_\theta(X_0))^2\, w(X_0)],$$

where w is a weight function to be chosen and $\mathbb{E}$ is the expectation $\mathbb{E}_{\theta^0,P_X}$. We assume that w is such that $(wf_\theta)^*/f_\varepsilon^*$ and $(wf_\theta^2)^*/f_\varepsilon^*$ are integrable, where φ^* is the Fourier transform of a function φ. We estimate θ^0 by $\widehat{\theta} = \arg\min_{\theta\in\Theta} S_n(\theta)$, where

$$S_n(\theta) = \frac{1}{2\pi n}\sum_{k=1}^n \mathbb{R}e \int \frac{\left(\left(Z_k - f_\theta\right)^2 w\right)^*(t)\, e^{-itZ_{k-1}}}{f_\varepsilon^*(-t)} dt, \tag{1.2}$$

where $\mathbb{R}e(u)$ is the real part of u. Under general assumptions, we prove that the estimator defined $\widehat{\theta}$ is consistent. Moreover, we give some conditions under which the parametric rate of convergence as well as the asymptotic normality can be stated. Those results hold under weak dependency conditions as introduced in Dedecker and Prieur (2005).

This procedure is clearly simpler than that of Comte and Taupin (2001). The resulting rate is more explicit and links directly the smoothness of the regression function to that of f_ε. Our new estimator is asymptotically Gaussian for a large class of regression functions, which is not the case in Comte and Taupin (2001).

The asymptotic properties of our estimator are illustrated through a simulation study. It confirms that our estimator performs well in various contexts, even in cases where the Markov chain $(X_i)_{i\geq 0}$ is not β-mixing (and not even irreducible), when the ratio signal to noise is small or large, for various sample sizes, and for different types of error distribution. Our estimator always better performs than the so-called naive estimator (built by replacing the non-observed X by Z in the usual least squares criterion). Our estimation procedure depends on the choice of the weight function w. The influence of this weight function is also studied in the simulations.

Finally, we propose a more general estimator when it is not possible to find a weight function w such that $(wf_\theta)^*/f_\varepsilon^*$ and $(wf_\theta^2)^*/f_\varepsilon^*$ are integrable. We establish a consistency result, and we give an upper bound for the quadratic risk, that relates the smoothness properties of the regression function to that of f_ε. These last results are proved under α-mixing conditions.

The paper is organized as follows. In Section 2 we present our estimation procedure. The theoretical properties of the estimator are stated in Section 3. The simulations are presented in Section 4. In Section 5 we introduce a more general estimator and we describe its asymptotic behavior. The proofs are gathered in Appendix.

2. Estimation procedure

In order to define more rigorously the criterion presented in the introduction, we first give some preliminary notations and assumptions.

2.1. **Notations.** Let

$$\| \varphi \|_1 = \int |\varphi(x)|dx, \ \| \varphi \|_2^2 = \int \varphi^2(x)dx, \ \text{ and } \ \| \varphi \|_\infty = \sup_{x\in\mathbb{R}} |\varphi(x)|.$$

The convolution product of two square integrable functions p and q is denoted by $p \star q(z) = \int p(z-x)q(x)dx$. The Fourier transform φ^* of a function φ is defined by

$$\varphi^*(t) = \int e^{itx}\varphi(x)dx.$$

For $\theta \in \mathbb{R}^d$, let $\| \theta \|_{\ell^2}^2 = \sum_{k=1}^d \theta_k^2$, and let $\theta^\top$ be the transpose matrix of θ.

For a map $(\theta, u) \mapsto \varphi_\theta(u)$ from $\Theta \times \mathbb{R}$ to $\mathbb{R}$, the first and second derivatives with respect to θ are denoted by

$$\begin{aligned} \varphi_\theta^{(1)}(\cdot) &= \left(\varphi_{\theta,j}^{(1)}(\cdot)\right)_{1\leq j\leq d}, \ \text{ with } \varphi_{\theta,j}^{(1)}(\cdot) = \frac{\partial \varphi_\theta(\cdot)}{\partial\theta_j} \text{ for } j \in \{1,\cdots,d\} \\ \text{and} \quad \varphi_\theta^{(2)}(\cdot) &= \left(\varphi_{\theta,j,k}^{(2)}(\cdot)\right)_{1\leq j,k\leq d}, \ \text{ with } \varphi_{\theta,j,k}^{(2)}(\cdot) = \frac{\partial^2 \varphi_\theta(\cdot)}{\partial\theta_j\partial\theta_k}, \text{ for } j,k \in \{1,\cdots,d\}. \end{aligned}$$

From now, $\mathbb{P}$, $\mathbb{E}$ and Var denote respectively the probability $\mathbb{P}_{\theta^0,P_X}$, the expected value $\mathbb{E}_{\theta^0,P_X}$ and the variance $\mathrm{Var}_{\theta^0,P_X}$, when the underlying and unknown true parameters are θ^0 and P_X.

2.2. **Assumptions.** We consider three types of assumptions.

• **Smoothness and moment assumptions**

($\mathbf{A_1}$) On Θ°, the function $\theta \mapsto f_\theta$ admits continuous derivatives with respect to θ up to the order 3.

($\mathbf{A_2}$) On Θ°, the quantity $w(X_0)(Z_1 - f_\theta(X_0))^2$, and the absolute values of its derivatives with respect to θ up to order 2 have a finite expectation.

• **Identifiability assumptions**

($\mathbf{I1_1}$) The quantity $S_{\theta^0,P_X}(\theta) = \mathbb{E}[(f_{\theta^0}(X) - f_\theta(X))^2 w(X)]$ admits one unique minimum at $\theta = \theta^0$.

($\mathbf{I1_2}$) For all $\theta \in \Theta^\circ$, the matrix $S^{(2)}_{\theta^0,P_X}(\theta) = \left(\dfrac{\partial^2 S_{\theta^0,P_X}(\theta)}{\partial\theta_i \partial\theta_j}\right)_{1\leq i,j\leq d}$ exists and the matrix

$$S^{(2)}_{\theta^0,P_X}(\theta^0) = 2\,\mathbb{E}\left[w(X)\left(f^{(1)}_{\theta^0}(X)\right)\left(f^{(1)}_{\theta^0}(X)\right)^\top\right]$$ is positive definite.

• **Assumptions on f_ε**

($\mathbf{N_1}$) The density f_ε belongs to $\mathbb{L}_2(\mathbb{R})$ and for all $x \in \mathbb{R}$, $f^*_\varepsilon(x) \neq 0$.

The assumption ($\mathbf{N_1}$) is quite usual when considering estimation in the convolution model. It ensures the existence of the estimation criterion.

2.3. **Definition of the estimator.** As already mentioned in the introduction, the starting point of our estimation procedure is to construct an estimator of the least square contrast

$$S_{\theta^0,P_X}(\theta) = \mathbb{E}[(Z_1 - f_\theta(X_0))^2\, w(X_0)], \tag{2.3}$$

based on the observations (Z_i) for $i = 0, \ldots, n$.

We consider the following condition: there exists a weight function w such that for all $\theta \in \Theta$,

($\mathbf{C_1}$) The functions (wf_θ) and (wf^2_θ) belong to $\mathbb{L}_1(\mathbb{R})$, and the functions w^*/f^*_ε, $(f_\theta w)^*/f^*_\varepsilon$, $(f^2_\theta w)^*/f^*_\varepsilon$ belong to $\mathbb{L}_1(\mathbb{R})$.

Remark 2.1. *The first part of Condition ($\mathbf{C_1}$) is not restrictive. The second part can be heuristically expressed as "one can find a weight function w such that wf_θ is smooth enough compared to f_ε". For a large number of regression functions, such a weight function can be easily exhibited. Some practical choices are discussed in the simulation study (Section 4).*

If ($\mathbf{C_1}$) holds, the expectations $\mathbb{E}(w(X))$, $\mathbb{E}(w(X)f_\theta(X))$ and $\mathbb{E}(w(X)f^2_\theta(X))$ can be easily estimated. Let us present the ideas of the estimation procedure. Let φ be such that φ and

$\varphi^*/f_\varepsilon^*$ belong to $\mathbb{L}_1(\mathbb{R})$. For such a function, due to the independence between ε_0 and X_0 we have

$$\mathbb{E}[\varphi(X_0)] = \mathbb{E}\left(\frac{1}{2\pi}\int \varphi^*(t)e^{-itX_0}dt\right) = \mathbb{E}\left(\frac{1}{2\pi}\int \frac{\varphi^*(t)e^{-itZ_0}}{f_\varepsilon^*(-t)}dt\right).$$

Hence, based on the observations $Z_0, \cdots, Z_n$, $\mathbb{E}[\varphi(X_0)]$ is estimated by

$$\frac{1}{2\pi}\mathbb{R}e\int \frac{\varphi^*(t)n^{-1}\sum_{j=1}^n e^{-itZ_j}}{f_\varepsilon^*(-t)}dt.$$

We then propose to estimate $S_{\theta^0,P_X}(\theta)$ by the quantity $S_n(\theta)$ defined by

$$S_n(\theta) = \frac{1}{2\pi n}\sum_{k=1}^n \mathbb{R}e\int \frac{\left((Z_k - f_\theta)^2 w\right)^*(t)\ e^{-itZ_{k-1}}}{f_\varepsilon^*(-t)}dt, \tag{2.4}$$

which satisfies

$$\mathbb{E}(S_n(\theta)) = \mathbb{E}[(Z_1 - f_\theta(X_0))^2 w(X_0)].$$

This criteria is minimum when $\theta = \theta^0$ under the identifiability assumption $(\mathbf{I1_1})$. Using this empirical criterion we propose to estimate θ^0 by

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} S_n(\theta). \tag{2.5}$$

3. Asymptotic properties

In this section, we give some conditions under which our estimator is consistent and asymptotically normal.

3.1. **Consistency of the estimator.**

The first result to mention is the consistency of our estimator. It holds under the following additional condition.

$(\mathbf{C_2})$ The functions $\sup_{\theta\in\Theta}\left|\left(f_{\theta,i}^{(1)}w\right)^*/f_\varepsilon^*\right|$ and $\sup_{\theta\in\Theta}\left|\left(f_\theta f_{\theta,i}^{(1)}w\right)^*/f_\varepsilon^*\right|$ belong to $\mathbb{L}_1(\mathbb{R})$ for any $i \in \{1, \dots, d\}$.

This condition is similar to $(\mathbf{C_1})$ for the first derivatives of f_θ. Thus it is not more restrictive than $(\mathbf{C_1})$.

Theorem 3.1. *Consider Model (1.1) under the assumptions $(\mathbf{A_1})$-$(\mathbf{A_2})$, $(\mathbf{I1_1})$, $(\mathbf{I1_2})$, $(\mathbf{N_1})$, and the conditions $(\mathbf{C_1})$-$(\mathbf{C_2})$. Then $\widehat{\theta}$ defined by (2.5) converges in probability to θ^0.*

3.2. $\sqrt{n}$**-consistency and asymptotic normality.** To state the asymptotic normality of our estimator, we need to introduce some additional conditions.

($\mathbf{C_3}$) the functions $\sup_{\theta\in\Theta}\left|\left(f^{(2)}_{\theta,i,j}w\right)^*/f^*_\varepsilon\right|$ and $\sup_{\theta\in\Theta}\left|\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}(f^2_\theta w)\right)^*/f^*_\varepsilon\right|$ belong to $\mathbb{L}_1(\mathbb{R})$ for any $i,j\in\{1,\dots,d\}$;

($\mathbf{C_4}$) the functions $\sup_{\theta\in\Theta}\left|\left(\frac{\partial^3(f_\theta w)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right)^*/f^*_\varepsilon\right|$ and $\sup_{\theta\in\Theta}\left|\left(\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}(f^2_\theta w)\right)^*/f^*_\varepsilon\right|$ belong to $\mathbb{L}_1(\mathbb{R})$, for $i,j,k\in\{1,\cdots,d\}$.

($\mathbf{C_5}$) The integrals $\int|t(f_{\theta^0}w)^*(t)|dt$ and $\int|t(f_{\theta^0}f^{(1)}_{\theta^0,k}w)^*(t)|dt$ are finite, for $k\in\{1,\cdots,d\}$.

The asymptotic properties of $\widehat{\theta}$, defined by (2.5), are stated under two different dependency conditions, which are presented below.

Definition 3.1. *Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space. Let Y be a random variable with values in a Banach space $(\mathbb{B},\|\cdot\|_{\mathbb{B}})$. Denote by $\Lambda_\kappa(\mathbb{B})$ the set of κ-Lipschitz functions, i.e. the functions f from $(\mathbb{B},\|\cdot\|_{\mathbb{B}})$ to $\mathbb{R}$ such that $|f(x)-f(y)|\le\kappa\|x-y\|_{\mathbb{B}}$. Let $\mathcal{M}$ be a σ-algebra of $\mathcal{A}$. Let $\mathbb{P}_{Y|\mathcal{M}}$ be a conditional distribution of Y given $\mathcal{M}$, $\mathbb{P}_Y$ the distribution of Y, and $\mathcal{B}(\mathbb{B})$ the Borel σ-algebra on $(\mathbb{B},\|\cdot\|_{\mathbb{B}})$. The dependence coefficients α and τ are defined by*

$$\begin{aligned}\alpha(\mathcal{M},\sigma(Y)) &= \frac{1}{2}\sup_{A\in\mathcal{B}(\mathbb{B})}\mathbb{E}(|\mathbb{P}_{Y|\mathcal{M}}(A)-\mathbb{P}_Y(A)|)\,,\\ \text{and if } \mathbb{E}(\|Y\|_{\mathbb{B}})<\infty,\quad \tau(\mathcal{M},Y) &= \mathbb{E}\Big(\sup_{f\in\Lambda_1(\mathbb{B})}|\mathbb{P}_{Y|\mathcal{M}}(f)-\mathbb{P}_Y(f)|\Big)\,.\end{aligned}$$

Let $\mathbf{X}=(X_i)_{i\ge0}$ be a strictly stationary Markov chain of real-valued random variables. On $\mathbb{R}^2$, we put the norm $\|x\|_{\mathbb{R}^2}=(|x_1|+|x_2|)/2$. For any integer $k\ge0$, the coefficients $\alpha_{\mathbf{X}}(k)$ and $\tau_{\mathbf{X},2}(k)$ of the chain are defined by

$$\begin{aligned}\alpha_{\mathbf{X}}(k) &= \alpha(\sigma(X_0),\sigma(X_k))\\ \text{and if } \mathbb{E}(|X_0|)<\infty,\quad \tau_{\mathbf{X},2}(k) &= \sup\{\tau(\sigma(X_0),(X_{i_1},X_{i_2})),k\le i_1\le i_2\}\,.\end{aligned}$$

Coefficient $\alpha(\mathcal{M},\sigma(Y))$ is the usual strong mixing coefficient introduced by Rosenblatt (1956). Coefficient $\tau(\mathcal{M},Y)$ has been introduced by Dedecker and Prieur (2005). In Section A.2, we recall some conditions on ξ_0 and f_{θ^0} under which the Markov chain $(X_i)_{i\ge0}$ is α-mixing or τ-dependent and illustrate those conditions through some examples.

First we state the asymptotic normality of $\widehat{\theta}$ when the Markov chain (X_i) of Model (1.1) is α-mixing.

Theorem 3.2. *Consider Model (1.1) under assumptions ($\mathbf{A_1}$), ($\mathbf{A_2}$), ($\mathbf{I1_1}$), ($\mathbf{I1_2}$), ($\mathbf{N_1}$), and conditions ($\mathbf{C_1}$)-($\mathbf{C_4}$). Let $Q_{|X_1|}$ be the inverse cadlag of the tail function $t\to\mathbb{P}(|X_1|>t)$. Assume that*

$$\sum_{k\ge1}\int_0^{\alpha_{\mathbf{X}}(k)}Q^2_{|X_1|}(u)du<\infty\,. \tag{3.6}$$

Then $\widehat{\theta}$ defined by (2.5) is a $\sqrt{n}$-consistent estimator of θ^0 which satisfies

$$\sqrt{n}(\widehat{\theta}-\theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma_1),$$

where the covariance matrix Σ_1 is defined in equation (B.5).

Next, we give the corresponding result when the Markov chain (X_i) is τ-dependent.

Theorem 3.3. *Consider Model (1.1) under assumptions ($\mathbf{A_1}$), ($\mathbf{A_2}$), ($\mathbf{I1_1}$), ($\mathbf{I1_2}$), ($\mathbf{N_1}$), and conditions ($\mathbf{C_1}$)-($\mathbf{C_5}$). Let $G(t) = t^{-1}\mathbb{E}(X_1^2\mathbf{1}_{X_1^2>t})$, and let G^{-1} be the inverse cadlag of G. Assume that*

$$\sum_{k>0} G^{-1}(\tau_{\mathbf{X},2}(k))\tau_{\mathbf{X},2}(k) < \infty\,. \tag{3.7}$$

Then $\widehat{\theta}$ defined by (2.5) is a $\sqrt{n}$-consistent estimator of θ^0 which satisfies

$$\sqrt{n}(\widehat{\theta}-\theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma_1),$$

where the covariance matrix Σ_1 is defined in equation (B.5).

Remark 3.1. *Let us give some conditions under which (3.6) or (3.7) are verified. Assume that $\mathbb{E}(|X_0|^p) < \infty$ for some $p > 2$. Then (3.6) is true provided that $\sum_{k>0} k^{2/(p-2)}\alpha_{\mathbf{X}}(k) < \infty$, and (3.7) is true provided that $\sum_{k>0}(\tau_{\mathbf{X},2}(k))^{(p-2)/p} < \infty$.*

Note that those results do not require the Markov chain to be absolutely regular as it is the case in Comte and Taupin (2001). Consequently they apply to autoregressive models with weaker dependency conditions. Beside the dependency conditions, our estimation procedure allows to achieve the parametric rate for a larger class of regression functions than in Comte and Taupin (2001).

The conditions under which Theorems 3.2 and 3.3 hold are similar, except Condition ($\mathbf{C_5}$) which appears only in Theorem 3.3. This condition is just technical and not restrictive at all.

The choice of the weight function w is crucial. Various weight functions can handle with Conditions $\mathbf{C_1}$-$\mathbf{C_5}$. The numerical properties of the resulting estimators will differ from one choice to another. This point is discussed on simulated data in the next section.

4. Simulation study

We investigate the properties of our estimator for different regression functions on simulated data. For each choice of regression function, we consider two error distributions: the Laplace distribution and the Gaussian distribution. When ε_1 has the Laplace distribution, its density and Fourier transform are

$$f_\varepsilon(x) = \frac{1}{\sigma_\varepsilon\sqrt{2}}\exp\Big(-\frac{\sqrt{2}}{\sigma_\varepsilon}|x|\Big), \text{ and } f_\varepsilon^*(x) = \frac{1}{1+\sigma_\varepsilon^2x^2/2}. \tag{4.8}$$

Hence, ε_1 is centered with variance σ_ε^2.

When ε_1 is Gaussian, its density and Fourier transform are

$$f_\varepsilon(x) = \frac{1}{\sigma_\varepsilon\sqrt{2\pi}}\exp\Big(-\frac{x^2}{2\sigma_\varepsilon^2}\Big), \text{ and } f_\varepsilon^*(x) = \exp(-\sigma_\varepsilon^2x^2/2). \tag{4.9}$$

Hence, ε_1 is centered with variance σ_ε^2.

For each of these error distributions, we consider the case of a linear regression function and of a Cauchy regression function. We start with the linear case.

4.1. **Linear regression function.** We consider the model (1.1) with $f_\theta(x) = ax+b$, where $|a| < 1$ and $\theta = (a,b)^T$. In these simulations, we have chosen to illustrate the numerical properties of our estimator under the weakest of the dependency conditions, that is τ-dependency. As it is recalled in Appendix A.2, when f_{θ^0} is linear with $|a| < 1$, if ξ_0 has a density bounded from below in a neighborhood of the origin, then the Markov chain $(X_i)_{i\geq 0}$ is α-mixing. When ξ_0 does not have a density, then the chain may not be α-mixing (and not even irreducible), but it is always τ-dependent.

Here, we consider the case where the innovation distribution is discrete, in such a way that the stationary Markov Chain is τ-dependent but not α-mixing. We also consider two distinct values of θ_0. For the first value, the stationary distribution of X_i is absolutely continuous with respect to the Lebesgue measure. For the second value, the stationary distribution is singular with respect to the Lebesgue measure. In both cases Theorem 3.3 applies, and the estimator $\hat{\theta}$ is asymptotically normal.

• Case A (absolutely continuous stationary distribution). We focus on the case where the true parameter is $\theta^0 = (1/2, 1/4)^T$, X_0 is uniformly distributed over $[0,1]$, and $(\xi_i)_{i\geq 1}$ is a sequence of i.i.d. random variables, independent of X_0 and such that $\mathbb{P}(\xi_1 = -1/4) = \mathbb{P}(\xi_1 = 1/4) = 1/2$. Then the Markov chain defined for $i > 0$ by

$$X_i = \frac{1}{4} + \frac{1}{2}X_{i-1} + \xi_i \tag{4.10}$$

is strictly stationary, the stationary distribution being the uniform distribution over $[0,1]$, and consequently $\sigma_{X_0}^2 = 1/12$. This chain is non-irreducible, and the dependency coefficients are such that $\alpha_{\mathbf{X}}(k) = 1/4$ (see for instance Bradley (1986), p. 180) and $\tau_{\mathbf{X},2}(k) = O(2^{-k})$. Thus the Markov chain is not α-mixing, but it is τ-dependent. For the simulation, we start with X_0 uniformly distributed over $[0,1]$, so the simulated chain is stationary.

• Case B (singular stationary distribution). We consider the case where the true parameter is $\theta^0 = (1/3, 1/3)^T$, X_0 is uniformly distributed over the Cantor set, and $(\xi_i)_{i\geq 1}$ is a sequence of i.i.d. random variables, independent of X_0 and such that $\mathbb{P}(\xi_1 = -1/3) = \mathbb{P}(\xi_1 = 1/3) = 1/2$. Then the Markov chain defined for $i > 0$ by

$$X_i = \frac{1}{3} + \frac{1}{3}X_{i-1} + \xi_i \tag{4.11}$$

is strictly stationary, the stationary distribution being the uniform distribution over the Cantor set, and consequently $\sigma_X^2 = 1/8$. This chain is non-irreducible, and the dependency coefficients satisfy $\alpha_{\mathbf{X}}(k) = 1/4$ and $\tau_{\mathbf{X},2}(k) = O(3^{-k})$. Thus the Markov chain is not α-mixing, but is τ-dependent. For the simulation, we start with X_0 uniformly distributed over $[0,1]$, and we consider that the chain is close to the stationary chain after 1000 iterations. We then set $X_i = X_{i+1000}$.

In these two cases, we can find a weight function w satisfying the conditions $(\mathbf{C_1})$-$(\mathbf{C_5})$. We first give the detailed expression of the estimator for two choices of weight functions w. Then we recall the classic estimator when X is directly observed, the ARMA estimator, and the so-called naive estimator.

4.1.1. *Expression of the estimator.* We consider the two following weight functions w

$$w(x) = N(x) = \exp\{-x^2/(4\sigma_\varepsilon^2)\} \quad \text{and} \quad w(x) = SC(x) = \frac{1}{2\pi}\Big(\frac{2 * \sin(x)}{x}\Big)^4. \tag{4.12}$$

These choices of weight ensure that Conditions $(\mathbf{C_1})$-$(\mathbf{C_5})$ hold and that the two estimators, denoted by $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$ respectively, converge to θ^0 with the parametric rate of convergence. There are two main differences between these two weight functions. First, N depends on the variance error σ_ε^2. Hence the estimator should be adaptive to the noise level. On the contrary, it may be sensitive to very small error variance as it appears in the simulations (see Figure 1). Second, SC has strong smoothness properties since its Fourier transform is compactly supported.

The two associated estimators are based on the calculation of $S_n(\theta)$, which can be written as

$$S_n(\theta) = \frac{1}{n}\sum_{k=1}^{n}[(Z_k^2 + b^2 - 2Z_k b)I_0(Z_{k-1}) + a^2 I_2(Z_{k-1}) - 2a(Z_k - b)I_1(Z_{k-1})],$$

with

$$I_j(Z) = \frac{1}{2\pi}\mathbb{R}e\int (p_j w)^*(u)\frac{e^{-iuZ}}{f_\varepsilon^*(-u)}du, \tag{4.13}$$

where $p_j(x) = x^j$ for $j = 0, 1, 2$, w being either $w = N$ or $w = SC$. With the above notations, $\widehat{\theta} = (\widehat{a}, \widehat{b})^T$ satisfies

$$\widehat{a} = \frac{\sum_{k=1}^n Z_k I_1(Z_{k-1})\sum_{k=1}^n I_0(Z_{k-1}) - \sum_{k=1}^n Z_k I_0(Z_{k-1})\sum_{k=1}^n I_1(Z_{k-1})}{\sum_{k=1}^n I_2(Z_{k-1})\sum_{k=1}^n I_0(Z_{k-1}) - \big(\sum_{k=1}^n I_1(Z_{k-1})\big)^2}, \tag{4.14}$$

$$\widehat{b} = \frac{\sum_{k=1}^n Z_k I_0(Z_{k-1})}{\sum_{k=1}^n I_0(Z_{k-1})} - \widehat{a}\frac{\sum_{k=1}^n I_1(Z_{k-1})}{\sum_{k=1}^n I_0(Z_{k-1})}. \tag{4.15}$$

We now compute $I_j(Z)$ for $j = 0, 1, 2$ and the two weight functions. In the following we respectively denote $I_{j,N}(Z)$ and $I_{j,SC}(Z)$ the previous integrals when the weight function is either $w = N$ or $w = SC$.

We start with $w = N$ and give the details of the calculations for the two error distributions (Laplace and Gaussian), which are explicit. Then, with the weight function $w = SC$, we present the calculations, which are not explicit whatever the error distribution f_ε.

- When $w = N$, Fourier calculations provide that

$$\begin{aligned}
N^*(t) &= \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2) \\
(Np_1)^*(t) &= \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(-2\sigma_\varepsilon^2 t/i\big), \\
(Np_2)^*(t) &= -\sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(-2\sigma_\varepsilon^2 + 4\sigma_\varepsilon^4 t^2\big).
\end{aligned}$$

It follows that

$$\begin{aligned}
I_{0,N}(Z) &= \frac{1}{2\pi}\mathbb{R}e\int \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\frac{e^{-itZ}}{f_\varepsilon^*(-t)}dt,\\
I_{1,N}(Z) &= \frac{1}{2\pi}\mathbb{R}e\int \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(-2\sigma_\varepsilon^2 t/i\big)\frac{e^{-itZ}}{f_\varepsilon^*(-t)}dt,\\
I_{2,N}(Z) &= \frac{1}{2\pi}\mathbb{R}e\int \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(2\sigma_\varepsilon^2-4\sigma_\varepsilon^4 t^2\big)\frac{e^{-itZ}}{f_\varepsilon^*(-t)}dt.
\end{aligned}$$

If f_ε is the Laplace distribution (4.8), replacing f_ε^* by its expression we get

$$I_{0,N}(Z) = e^{-Z^2/(4\sigma_\varepsilon^2)} - \frac{\sigma_\varepsilon^2}{2}\frac{\partial^2}{\partial Z^2}N(Z) = \left[5/4 - Z^2/(8\sigma_\varepsilon^2)\right]e^{-Z^2/(4\sigma_\varepsilon^2)},$$

$$I_{1,N}(Z) = \left[7Z/4 - Z^3/(8\sigma_\varepsilon^2)\right]e^{-Z^2/(4\sigma_\varepsilon^2)}, I_{2,N}(Z) = \left[-\sigma_\varepsilon^2 + 9Z^2/4 - Z^4/(8\sigma_\varepsilon^2)\right]e^{-Z^2/(4\sigma_\varepsilon^2)}.$$

If f_ε is the Gaussian distribution (4.9), replacing f_ε^* by its expression we obtain $I_{0,N}(Z) = \sqrt{2}e^{-Z^2/(2\sigma_\varepsilon^2)}$, $I_{1,N}(Z) = 2\sqrt{2}Ze^{-Z^2/(2\sigma_\varepsilon^2)}$ and $I_{2,N}(Z) = \sqrt{2}(4Z^2-2\sigma_\varepsilon^2)e^{-Z^2/(2\sigma_\varepsilon^2)}$.

Hence we deduce the expression of $\widehat{a}_N$ and $\widehat{b}_N$ by applying (4.14) and (4.15).

•When $w = SC$, Fourier calculations provide that

$$\begin{aligned}
SC^*(t) &= \mathbb{1}_{[-4,-2]}(t)(t^3/6+2t^2+8t+32/3) + \mathbb{1}_{[-2,0]}(t)(-t^3/2-2t^2+16/3)\\
&\quad +\mathbb{1}_{[2,4]}(t)(-t^3/6+2t^2-8t+32/3) + \mathbb{1}_{[0,2]}(t)(t^3/2-2t^2+16/3)\\
(SCp_1)^*(t) &= \frac{\partial}{\partial t}SC^*(t)/i \text{ and } (SCp_2)^*(t) = \frac{\partial^2}{\partial t^2}SC^*(t)/(i^2).
\end{aligned}$$

The integrals $I_{j,SC}(Z)$, defined for $j = 0,1,2$ by

$$I_{j,SC}(Z) = \frac{1}{2\pi}\mathbb{R}e\int (SCp_j)^*(t)\frac{e^{-itZ}}{f_\varepsilon^*(-t)}dt, \tag{4.16}$$

have no explicit form, whatever the error distribution f_ε. It has to be numerically computed, using the IFFT Matlab function. More precisely, we consider a finite Fourier series approximation of $(SCp_j)^*(t)/f_\varepsilon^*(t)$ whose Fourier transfom is calculated using IFFT Matlab function. The result is taken as an approximation of $I_{j,SC}(Z)$. Finally we deduce the expression of $\widehat{a}_{SC}$ and $\widehat{b}_{SC}$ by applying (4.14) and (4.15).

4.1.2. *Comparison with classical estimators.* We compare the two estimators $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$ with three classical estimators, the usual least square estimator when there is no observation noise, the ARMA estimator, and the so-called naive estimator.

• Estimator without noise. In the case where $\varepsilon_i = 0$, that is $(X_0,\dots,X_n)$ is observed without error, the parameters can be easily estimated by the usual least square estimators

$$\widehat{a}_X = \frac{n\sum_{i=1}^n X_iX_{i-1} - \sum_{i=1}^n X_i\sum_{i=1}^n X_{i-1}}{n\sum_{i=1}^n X_{i-1}^2 - (\sum_{i=1}^n X_{i-1})^2} \quad\text{and}\quad \widehat{b}_X = \frac{1}{n}\Big(\sum_{i=1}^n X_i\Big) - \widehat{a}_X\frac{1}{n}\Big(\sum_{i=1}^n X_{i-1}\Big).$$

• ARMA estimator. When the regression function is linear, the model may be written as

$$Z_i - aZ_{i-1} - b = \xi_i + \varepsilon_i - a\varepsilon_{i-1}\,.$$

The auto-covariance function γ_Y of the stationary sequence $Y_i = \xi_i + \varepsilon_i - a\varepsilon_{i-1}$ is given by

$$\gamma_Y(0) = (1+a^2)\sigma_\varepsilon^2 + \sigma_\xi^2, \quad \gamma_Y(1) = -a\sigma_\varepsilon^2, \quad \text{and } \gamma_Y(k) = 0 \text{ for } k > 1.$$

It follows that Y_i is an MA(1) process, which may be written as

$$Y_i = \eta_i - \beta\eta_{i-1},$$

where η_i is the innovation, and $|\beta| < 1$ (note that $|\beta| \neq 1$ because $\gamma_Y(0) - 2|\gamma_Y(1)| > 0$). Moreover, one can give the explicit expression of β and σ_η^2 in terms of a, σ_ξ^2 and σ_ε^2. It follows that, if $|a| < 1$, $(Z_i)_{i\geq 0}$ is the causal invertible ARMA(1,1) process

$$Z_i - aZ_{i-1} = b + \eta_i - \beta\eta_{i-1}. \tag{4.17}$$

Note that $a \neq \beta$ except if $a = 0$. Hence, if $|a| < 1$ and $a \neq 0$, one can estimate the parameters (a, b, β) by maximizing the so-called Gaussian likelihood. These estimators are consistent and asymptotically Gaussian. Moreover they are efficient when both the innovations and the errors ε are Gaussian (see Hannan (1973) or Brockwell and Davis (1991)). Note that this well-known approach does not require the knowledge of the error distribution, but of course it works only in the particular case where the regression function f_θ is linear. For the computation of the ARMA estimator we use the function *arma* from the R *tseries* package (see Trapletti and Hornik (2011)). The resulting estimators are denoted by $\widehat{a}_{arma}$ and $\widehat{b}_{arma}$.

• Naive estimator. The naive estimator is constructed by replacing the unobserved X_i by the observation Z_i in the expression of $\widehat{a}_X$ and $\widehat{b}_X$:

$$\widehat{a}_{naive} = \frac{n\sum_{i=1}^n Z_i Z_{i-1} - \sum_{i=1}^n Z_i \sum_{i=1}^n Z_{i-1}}{n\sum_{i=1}^n Z_{i-1}^2 - (\sum_{i=1}^n Z_{i-1})^2} \quad \text{and} \quad \widehat{b}_{naive} = \frac{1}{n}\Big(\sum_{i=1}^n Z_i\Big) - \widehat{a}_{naive}\frac{1}{n}\Big(\sum_{i=1}^n Z_{i-1}\Big).$$

Classical results show that $\widehat{\theta}_{naive}$ is an asymptotically biased estimator of θ^0, which is confirmed by the simulation study.

4.1.3. *Simulation results.* For each error distribution, we simulate 100 samples with size n, $n = 500$, 5000 and 10000. We consider different values of σ_ε such that the ratio signal to noise $s2n = \sigma_\varepsilon^2/\mathrm{Var}(X)$ is $0.5, 1.5$ or 3. The comparison of the five estimators is based on the bias, the Mean Squared Error (MSE), and the box plots. If $\widehat{\theta}(k)$ denotes the value of the estimation for the k-th sample, the MSE is evaluated by the empirical mean over the 100 samples:

$$MSE(\widehat{\theta}) = \frac{1}{100}\sum_{k=1}^{100}(\widehat{\theta}(k) - \theta^0)^2.$$

Results are presented in Figures 1-2 and Tables 1-4.

The first thing to notice is that, not surprisingly, $\widehat{\theta}_{naive}$ presents a bias, whatever the values of n, $s2n$ and the error distribution. The estimator $\widehat{\theta}_X$ has the good expected properties (unbiased and small MSE), but it is based on the observation of the X_i's. The previously known estimator

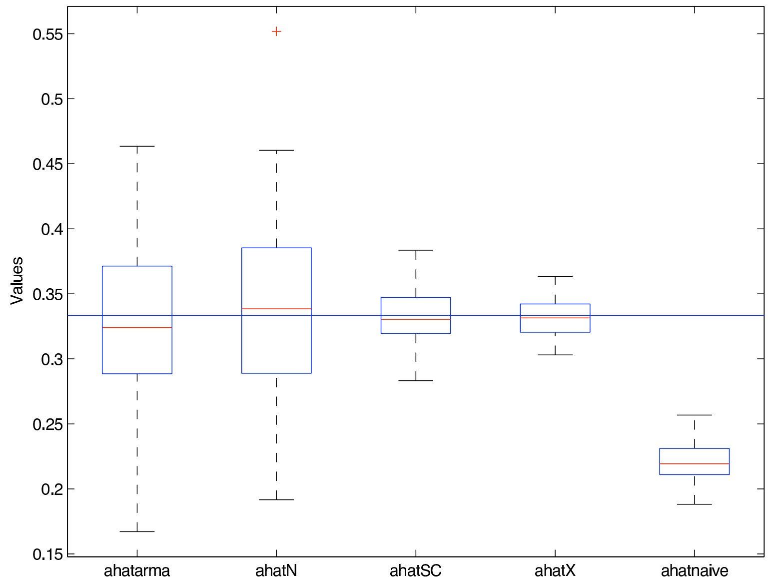

FIGURE 1. Results for linear Case B and Gaussian error, with $n = 5000$ and $\sigma_\varepsilon^2/\mathrm{Var}(X) = 0.5$. Box plots of the five estimators $\widehat{a}_{arma}$, $\widehat{a}_N$, $\widehat{a}_{SC}$, $\widehat{a}_X$ and $\widehat{a}_{naive}$, from left to right, based on 100 replications. True value is 1/3 (horizontal line).

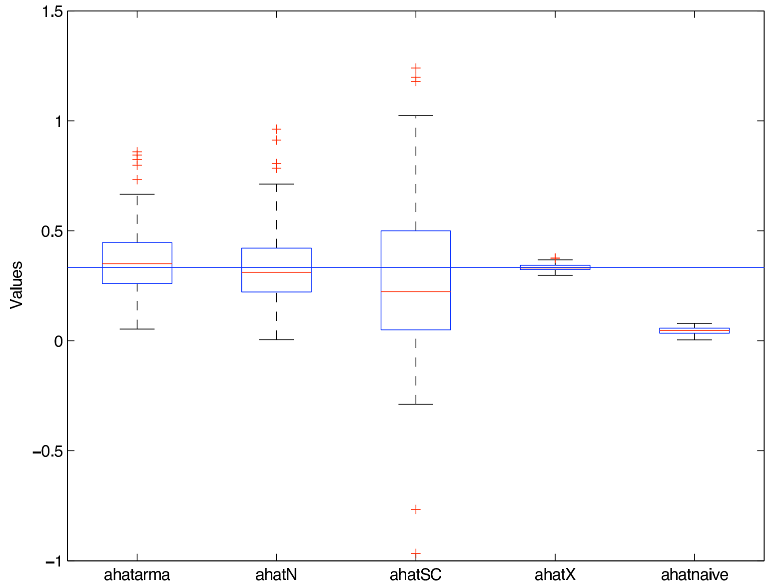

FIGURE 2. Results for linear Case B and Gaussian error, with $n = 5000$ and $\sigma_\varepsilon^2/\mathrm{Var}(X) = 6$. Box plots of the five estimators $\widehat{a}_{arma}$, $\widehat{a}_N$, $\widehat{a}_{SC}$, $\widehat{a}_X$ and $\widehat{a}_{naive}$, from left to right, based on 100 replications. True value is 1/3 (horizontal line).

	ratio		Estimator				
n	s2n		$\widehat{\theta}_{arma}(MSE)$	$\widehat{\theta}_N(MSE)$	$\widehat{\theta}_{SC}(MSE)$	$\widehat{\theta}_X(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	a	0.487 (0.008)	0.459 (0.020)	0.489 (0.002)	0.493 (0.001)	0.328 (0.030)
		b	0.257 (0.002)	0.262 (0.002)	0.255 (0.001)	0.253 (0.001)	0.336 (0.008)
	1.5	a	0.494 (0.015)	0.488 (0.013)	0.492 (0.006)	0.501 (0.001)	0.198 (0.092)
		b	0.251 (0.004)	0.253 (0.002)	0.253 (0.002)	0.249 (0.001)	0.399 (0.023)
	3	a	0.461 (0.044)	0.502 (0.029)	0.503 (0.026)	0.493 (0.001)	0.121 (0.145)
		b	0.270 (0.012)	0.249 (0.001)	0.249 (0.001)	0.253 (0.001)	0.440 (0.037)
5000	0.5	a	0.497 (0.001)	0.499 (0.004)	0.499 (0.001)	0.499 (0.001)	0.332 (0.028)
		b	0.252 (0.001)	0.251 (0.001)	0.251 (0.001)	0.251 (0.001)	0.334 (0.007)
	1.5	a	0.498 (0.003)	0.508 (0.003)	0.503 (0.002)	0.499 (0.001)	0.199 (0.091)
		b	0.250 (0.001)	0.247 (0.001)	0.248 (0.001)	0.250 (0.001)	0.399 (0.022)
	3	a	0.487 (0.008)	0.492 (0.004)	0.495 (0.004)	0.500 (0.001)	0.123 (0.143)
		b	0.256 (0.002)	0.253 (0.001)	0.252 (0.001)	0.250 (0.001)	0.437 (0.035)
10000	0.5	a	0.496 (0.001)	0.501 (0.002)	0.500 (0.001)	0.499 (0.001)	0.334 (0.028)
		b	0.252 (0.001)	0.250 (0.001)	0.250 (0.001)	0.250 (0.001)	0.333 (0.007)
	1.5	a	0.504 (0.002)	0.500 (0.001)	0.501 (0.001)	0.500 (0.001)	0.200 (0.090)
		b	0.248 (0.001)	0.250 (0.001)	0.250 (0.001)	0.250 (0.001)	0.401 (0.023)
	3	a	0.493 (0.003)	0.499 (0.001)	0.499 (0.002)	0.498 (0.001)	0.124 (0.142)
		b	0.254 (0.001)	0.250 (0.001)	0.250 (0.001)	0.251 (0.001)	0.438 (0.036)

TABLE 1. Estimation results for Linear Case A, Laplace error. Mean estimated values of the five estimators $\widehat{\theta}_{arma}$, $\widehat{\theta}_N$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/2$, $b^0 = 1/4$. MSEs are given in brackets.

$\widehat{\theta}_{arma}$ has good asymptotic properties. However its bias is often larger than the biases of $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$, except when $s2n = 0.5$ and ε is Gaussian.

We now consider the two estimators $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$. Recall that their construction requires the choice of w. Note first that, whatever the weight function w, the two estimators $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$ present good convergence properties. Their biases and MSEs decrease when n increases. When compared one to another, we can see that their numerical behaviors are not the same. Namely for not too large $s2n$, $\widehat{\theta}_{SC}$ has a MSE smaller than $\widehat{\theta}_N$ (see Figure 1 and Tables 1-4, when $s2n \leq 3$). With large $s2n$, the estimator $\widehat{\theta}_N$ seems to have better properties (see Figure 2 when $s2n = 6$). This is expected since N depends on σ_ε^2 and is thus more sensitive to small values of σ_ε^2. The error distribution seems to have a slight infuence on the MSEs of the two estimators. The MSEs are often smaller when f_ε is the Laplace density. This may be related with the theoretical properties in density deconvolution. In that context it is well known that the rate of convergence is slower when f_ε is the Gaussian density. The two estimators $\widehat{\theta}_N$ and $\widehat{\theta}_{SC}$ have comparable numerical behaviors in the two linear autoregressive models. Let us recall that in both cases, the simulated chain X are non-mixing but are τ-dependent. In Case A, the stationary distribution of X is continuous whereas it is not the case in Case B. This explains

n	ratio s2n		Estimator $\widehat{\theta}_{arma}(MSE)$	$\widehat{\theta}_N(MSE)$	$\widehat{\theta}_{SC}(MSE)$	$\widehat{\theta}_X(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	a	0.483 (0.006)	0.539 (0.039)	0.496 (0.002)	0.495 (0.001)	0.331 (0.030)
		b	0.259 (0.002)	0.243 (0.003)	0.253 (0.001)	0.253 (0.001)	0.336 (0.008)
	1.5	a	0.497 (0.021)	0.516 (0.027)	0.507 (0.009)	0.499 (0.001)	0.200 (0.091)
		b	0.251 (0.005)	0.243 (0.005)	0.246 (0.002)	0.249 (0.001)	0.399 (0.023)
	3	a	0.456 (0.031)	0.521 (0.082)	0.481 (0.030)	0.501 (0.001)	0.120 (0.145)
		b	0.272 (0.008)	0.244 (0.016)	0.260 (0.007)	0.250 (0.001)	0.441 (0.037)
5000	0.5	a	0.497 (0.001)	0.492 (0.006)	0.499 (0.001)	0.498 (0.001)	0.333 (0.028)
		b	0.251 (0.001)	0.252 (0.001)	0.250 (0.001)	0.250 (0.001)	0.333 (0.007)
	1.5	a	0.490 (0.002)	0.510 (0.006)	0.502 (0.001)	0.499 (0.001)	0.120 (0.090)
		b	0.254 (0.001)	0.245 (0.001)	0.248 (0.001)	0.250 (0.001)	0.399 (0.022)
	3	a	0.471 (0.010)	0.512 (0.008)	0.503 (0.005)	0.498 (0.001)	0.124 (0.141)
		b	0.263 (0.002)	0.245 (0.002)	0.249 (0.001)	0.251 (0.001)	0.437 (0.035)
10000	0.5	a	0.504 (0.006)	0.500 (0.003)	0.498 (0.001)	0.499 (0.001)	0.331 (0.028)
		b	0.249 (0.001)	0.250 (0.001)	0.251 (0.001)	0.251 (0.001)	0.335 (0.007)
	1.5	a	0.495 (0.002)	0.501 (0.002)	0.499 (0.001)	0.501 (0.001)	0.200 (0.090)
		b	0.253 (0.001)	0.250 (0.001)	0.251 (0.001)	0.250 (0.001)	0.401 (0.023)
	3	a	0.492 (0.004)	0.498 (0.004)	0.500 (0.003)	0.500 (0.001)	0.126 (0.140)
		b	0.254 (0.001)	0.251 (0.001)	0.251 (0.001)	0.250 (0.001)	0.437 (0.009)

TABLE 2. Estimation results for Linear Case A, Gaussian error. Mean estimated values of the five estimators $\widehat{\theta}_{arma}$, $\widehat{\theta}_N$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/2$, $b^0 = 1/4$. MSEs are given in brackets.

the relative bad properties of $\widehat{\theta}_{arma}$ in Case B. Indeed, due to its construction, this estimator is expected to have good properties when the stationary distribution of the Markov Chain is close to the Gaussian distribution. On the contrary our estimators have similar behavior in both cases.

4.2. **Cauchy regression model.** We consider the model (1.1) with $f_\theta(x) = \theta/(1+x^2) = \theta f(x)$. The true parameter is $\theta^0 = 1.5$. For the law of ξ_0 we take $\xi_0 \sim \mathcal{N}(0, 0.01)$. In this case, an empirical study shows that σ_X^2 is about 0.1. Moreover $\alpha_{\mathbf{X}}(k) = O(\kappa^k)$ for some $\kappa \in]0,1[$ and the Markov chain is α-mixing (see Appendix A.2). For w suitably chosen, Theorem 3.2 applies and states that $\hat{\theta}$ is asymptotically normal. For the simulation, we start with X_0 uniformly distributed over $[0,1]$, and we consider that the chain is close to the stationary chain after 1000 iterations. We then set $X_i = X_{i+1000}$.

To our knowledge, the estimator $\widehat{\theta}$ is the first consistent estimator in the literature for this regression function. We first detail the estimator for two choices of the weight function w. Then we recall the classic estimator when X is directly observed and the so-called naive estimator.

n	ratio s2n		Estimator $\widehat{\theta}_{arma}(MSE)$	$\widehat{\theta}_N(MSE)$	$\widehat{\theta}_{SC}(MSE)$	$\widehat{\theta}_X(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	a	0.288 (0.021)	0.341 (0.013)	0.330 (0.002)	0.326 (0.001)	0.217 (0.015)
		b	0.354 (0.005)	0.331 (0.001)	0.333 (0.001)	0.335 (0.001)	0.389 (0.004)
	1.5	a	0.298 (0.050)	0.332 (0.009)	0.335 (0.007)	0.330 (0.001)	0.136 (0.040)
		b	0.349 (0.012)	0.331 (0.002)	0.329 (0.002)	0.335 (0.001)	0.429 (0.010)
	3	a	0.240 (0.127)	0.343 (0.017)	0.343 (0.018)	0.330 (0.001)	0.084 (0.063)
		b	0.385 (0.033)	0.333 (0.003)	0.333 (0.003)	0.338 (0.001)	0.465 (0.018)
5000	0.5	a	0.333 (0.004)	0.335 (0.003)	0.335 (0.001)	0.333 (0.001)	0.223 (0.012)
		b	0.333 (0.001)	0.332 (0.001)	0.332 (0.001)	0.334 (0.001)	0.388 (0.003)
	1.5	a	0.331 (0.011)	0.328 (0.002)	0.334 (0.001)	0.334 (0.001)	0.433 (0.041)
		b	0.334 (0.003)	0.334 (0.001)	0.329 (0.001)	0.332 (0.001)	0.132 (0.010)
	3	a	0.290 (0.030)	0.329 (0.003)	0.329 (0.004)	0.333 (0.001)	0.083 (0.063)
		b	0.355 (0.008)	0.335 (0.008)	0.335 (0.008)	0.334 (0.001)	0.459 (0.016)
10000	0.5	a	0.337 (0.002)	0.335 (0.002)	0.334 (0.001)	0.334 (0.001)	0.222 (0.012)
		b	0.331 (0.001)	0.332 (0.001)	0.332 (0.001)	0.332 (0.001)	0.388 (0.003)
	1.5	a	0.322 (0.006)	0.336 (0.001)	0.336 (0.001)	0.334 (0.001)	0.134 (0.040)
		b	0.339 (0.002)	0.332 (0.001)	0.332 (0.001)	0.333 (0.001)	0.433 (0.010)
	3	a	0.329 (0.010)	0.336 (0.002)	0.336 (0.002)	0.334 (0.001)	0.083 (0.063)
		b	0.335 (0.002)	0.332 (0.001)	0.332 (0.001)	0.332 (0.001)	0.457 (0.015)

TABLE 3. Estimation results for Linear Case B, Laplace error. Mean estimated values of the five estimators $\widehat{\theta}_{arma}$, $\widehat{\theta}_N$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/3$, $b^0 = 1/3$. MSEs are given in brackets.

4.2.1. *Expression of the estimator.* We consider the two following weight functions:

$$N_c(x) = (1+x^2)^2 \exp\{-x^2/(4\sigma_\varepsilon^2)\} \text{ and } SC_c(x) = (1+x^2)^2 \frac{1}{2\pi}\Big(\frac{2*\sin(x)}{x}\Big)^4, \tag{4.18}$$

with σ_ε^2 the variance of ε. This choice of w ensures that Conditions $(\mathbf{C_1})$-$(\mathbf{C_5})$ hold and our method allows to achieve the parametric rate of convergence. As in the linear case, these two weight functions differ by their dependence on σ_ε^2 and their smoothness properties. The two associated estimators are based on the calculation of $S_n(\theta)$, which can be written as

$$S_n(\theta) = \frac{1}{n}\sum_{k=1}^{n}[Z_k^2 I_w(Z_{k-1}) + \theta^2 I_{wf^2}(Z_{k-1}) - 2\theta Z_k I_{wf}(Z_{k-1})],$$

where

$$I_w(Z) = \frac{1}{2\pi}\mathbb{R}e\int (w)^*(u)\frac{e^{-iuZ}}{f_\varepsilon^*(-u)}du, \qquad I_{wf}(Z) = \frac{1}{2\pi}\mathbb{R}e\int (wf)^*(u)\frac{e^{-iuZ}}{f_\varepsilon^*(-u)}du$$

$$\text{and} \quad I_{wf^2}(Z) = \frac{1}{2\pi}\mathbb{R}e\int (wf^2)^*(u)\frac{e^{-iuZ}}{f_\varepsilon^*(-u)}du.$$

	ratio		Estimator				
n	s2n		$\widehat{\theta}_{arma}(MSE)$	$\widehat{\theta}_N(MSE)$	$\widehat{\theta}_{SC}(MSE)$	$\widehat{\theta}_X(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	a	0.327 (0.016)	0.349 (0.035)	0.330 (0.003)	0.326 (0.001)	0.218 (0.014)
		b	0.338 (0.004)	0.332 (0.002)	0.336 (0.001)	0.337 (0.001)	0.392 (0.004)
	1.5	a	0.290 (0.061)	0.355 (0.021)	0.345 (0.008)	0.332 (0.001)	0.133 (0.041)
		b	0.353 (0.015)	0.324 (0.004)	0.328 (0.002)	0.333 (0.001)	0.432 (0.010)
	3	a	0.234 (0.153)	0.329 (0.049)	0.329 (0.051)	0.326 (0.001)	0.077 (0.067)
		b	0.383 (0.040)	0.337 (0.010)	0.337 (0.010)	0.337 (0.001)	0.461 (0.017)
5000	0.5	a	0.329 (0.004)	0.341 (0.005)	0.333 (0.001)	0.332 (0.001)	0.220 (0.013)
		b	0.335 (0.001)	0.332 (0.001)	0.334 (0.001)	0.334 (0.001)	0.399 (0.003)
	1.5	a	0.329 (0.009)	0.331 (0.003)	0.332 (0.002)	0.333 (0.001)	0.132 (0.041)
		b	0.335 (0.002)	0.334 (0.001)	0.333 (0.001)	0.333 (0.001)	0.433 (0.010)
	3	a	0.315 (0.022)	0.348 (0.008)	0.348 (0.008)	0.334 (0.001)	0.084 (0.062)
		b	0.343 (0.006)	0.327 (0.002)	0.328 (0.002)	0.332 (0.001)	0.459 (0.016)
10000	0.5	a	0.330 (0.002)	0.333 (0.003)	0.333 (0.001)	0.332 (0.001)	0.221 (0.013)
		b	0.335 (0.001)	0.333 (0.001)	0.333 (0.001)	0.334 (0.001)	0.389 (0.003)
	1.5	a	0.328 (0.006)	0.336 (0.002)	0.334 (0.001)	0.333 (0.001)	0.132 (0.041)
		b	0.336 (0.002)	0.333 (0.001)	0.334 (0.001)	0.334 (0.001)	0.435 (0.010)
	3	a	0.312 (0.014)	0.334 (0.004)	0.334 (0.004)	0.333 (0.001)	0.083 (0.063)
		b	0.344 (0.003)	0.333 (0.001)	0.333 (0.001)	0.333 (0.001)	0.458 (0.016)

TABLE 4. Estimation results for Linear Case B, Gaussian error. Mean estimated values of the five estimators $\widehat{\theta}_{arma}$, $\widehat{\theta}_N$, $\widehat{\theta}_{SC}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True values are $a^0 = 1/3$, $b^0 = 1/3$. MSEs are given in brackets.

The estimator can be expressed as

$$\widehat{\theta} = \frac{\sum_{k=1}^n Z_k I_{wf}(Z_{k-1})}{\sum_{k=1}^n I_{wf^2}(Z_{k-1})}. \tag{4.19}$$

In the following we denote by $I_{wf,N_c}(Z)$, $I_{wf^2,N_c}(Z)$, $I_{wf,SC_c}(Z)$ and $I_{wf^2,SC_c}(Z)$ respectively, the previous integrals when the weight function is either $w = N_c$ or $w = SC_c$. In the same way we denote by $\widehat{\theta}_{N_c}$ and $\widehat{\theta}_{SC_c}$ the corresponding estimators of θ^0.

- When $w = N_c$, Fourier calculations provide that

$$\begin{aligned} (N_c f)^*(t) &= \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2)\big(1 + 2\sigma_\varepsilon^2(1 - 2\sigma_\varepsilon^2 t^2)\big) \\ \text{and } (N_c f^2)^*(t) &= \sqrt{2\pi}\sqrt{2\sigma_\varepsilon^2}\exp(-\sigma_\varepsilon^2 t^2). \end{aligned}$$

Now, we can calculate the integrals $I_{wf,N_c}(Z)$ and $I_{wf^2,N_c}(Z)$.

If f_ε is the Laplace distribution (4.8), replacing f_ε^* by its expression we obtain

$$\begin{aligned} I_{wf,N_c}(Z) &= \exp(-Z^2/(4\sigma_\varepsilon^2))\left[Z^4 - 18Z^2\sigma_\varepsilon^2 + Z^2 + 8\sigma_\varepsilon^4 - 10\sigma_\varepsilon^2\right]/(8\sigma_\varepsilon^2), \\ \text{and } I_{wf^2,N_c}(Z) &= \exp(-Z^2/(4\sigma_\varepsilon^2))\big[1 + \frac{1}{4}\big(1 - \frac{Z^2}{2\sigma_\varepsilon^2}\big)\big]. \end{aligned}$$

If f_ε is the Gaussian distribution (4.9), replacing f_ε^* by its expression we obtain

$$I_{wf,N_c}(Z) = \sqrt{2}e^{-Z^2/(2\sigma_\varepsilon^2)}(1 - 2\sigma_\varepsilon^2 + 4Z^2), \text{ and } \quad I_{wf^2,N_c}(Z) = \sqrt{2}e^{-Z^2/(2\sigma_\varepsilon^2)}.$$

• When $w = SC_c$, easy calculations show that

$$I_{wf,SC_c}(Z) = I_{0,SC}(Z) + I_{2,SC}(Z) \text{ and } I_{wf^2,SC_c}(Z) = I_{0,SC}(Z),$$

where $I_{0,SC}(Z)$ and $I_{2,SC}(Z)$ are defined by (4.16). As explained before, the integrals $I_{0,SC}(Z)$ and $I_{2,SC}(Z)$ have no explicit form, whatever the error distributions, and are numerically approximated via the IFFT function.

4.2.2. *Comparison with classical estimators.* We compare our estimators with two classical estimators, the usual least square estimator without observation noise, and the naive estimator.

• Estimator without noise. When $\varepsilon_i = 0$, that is $(X_0, \dots, X_n)$ is observed without errors, the parameter can be easily estimated by the usual least square estimator

$$\widehat{\theta}_X = \frac{\sum_{i=1}^n X_i f(X_{i-1})}{\sum_{i=1}^n f^2(X_{i-1})}.$$

• Naive estimator. The idea for the construction of the naive estimator is to replace the unobserved X_i by the observation Z_i in the expression of $\widehat{\theta}_X$ to get

$$\widehat{\theta}_{naive} = \frac{\sum_{i=1}^n Z_i f(Z_{i-1})}{\sum_{i=1}^n f^2(Z_{i-1})}.$$

Classical results show that $\widehat{\theta}_{naive}$ is an asymptotically biased estimator of θ^0, which is confirmed by the simulation study.

4.2.3. *Simulations results.* For each error distribution, we simulate 100 samples with size n, $n = 500$, 5000 and 10000. We consider different values of σ_ε such that the ratio signal to noise $s2n = \sigma_\varepsilon^2/\mathrm{Var}(X)$ is $0.5, 1.5$ or 3.

The comparison of the four estimators is based on the bias, the Mean Squared Error (MSE), and the box plots. The results are presented in Figure 3 and Tables 5-6.

The first thing to notice is that, not surprisingly, $\widehat{\theta}_{naive}$ presents a bias, whatever the values of n, $s2n$ and the errors distribution. Moreover it converges to (false) values which are different according to $s2n$ (see Tables (5)-(6)).

The estimator $\widehat{\theta}_X$ has the good expected properties (unbiased and small MSE), but it is based on the observation of the X_i's.

We now compare our two estimators illustrating the influence of w, $s2n$ and f_ε. Globally, whatever the weight function w, the two estimators $\widehat{\theta}$ present good convergence properties. Their biases and MSEs decrease when n increases. The MSEs of $\widehat{\theta}_{SC_c}$ increase when $s2n$ increases. This is not the case for the MSE of $\widehat{\theta}_{N_c}$. This is probably due to the fact that the weight

n	ratio s2n	Estimator $\widehat{\theta}_{N_c}(MSE)$	$\widehat{\theta}_{SC_c}(MSE)$	$\widehat{\theta}_X(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	1.5095 (0.0042)	1.5024 (0.0006)	1.5004 (0.0000)	1.4333 (0.0050)
	1.5	1.5006 (0.0021)	1.5005 (0.0013)	1.5002 (0.0000)	1.3657 (0.0190)
	3	1.5017 (0.0024)	1.5005 (0.0024)	1.5002 (0.0000)	1.3267 (0.0314)
5000	0.5	1.5045 (0.0008)	1.5005 (0.0001)	1.5003 (0.0000)	1.4320 (0.0047)
	1.5	1.5003 (0.0004)	1.4994 (0.0003)	1.4997 (0.0000)	1.3647 (0.0185)
	3	1.4989 (0.0005)	1.4992 (0.0005)	1.5000 (0.0000)	1.3223 (0.0318)
10000	0.5	1.5033 (0.0004)	1.5002 (0.0001)	1.5000 (0.0000)	1.4315 (0.0047)
	1.5	1.5000 (0.0002)	1.5000 (0.0001)	1.4998 (0.0000)	1.3650 (0.0183)
	3	1.4972 (0.0002)	1.4970 (0.0002)	1.4998 (0.0000)	1.3222 (0.0317)

TABLE 5. Estimation results for Cauchy, Laplace error. Mean estimated values of the four estimators $\widehat{\theta}_{N_c}$, $\widehat{\theta}_{SC_c}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True value is $\theta^0 = 1.5$. MSE are given in brackets.

n	ratio s2n	Estimator $\widehat{\theta}_{N_c}(MSE)$	$\widehat{\theta}_{SC_c}(MSE)$	$\widehat{\theta}_X(MSE)$	$\widehat{\theta}_{naive}(MSE)$
1000	0.5	1.4979 (0.0027)	1.4998 (0.0006)	1.5000 (0.0000)	1.4230 (0.0064)
	1.5	1.4995 (0.0029)	1.5001 (0.0015)	1.5005 (0.0000)	1.3336 (0.0287)
	3	1.5080 (0.0049)	1.5058 (0.0042)	1.4997 (0.0000)	1.2832 (0.0487)
5000	0.5	1.5033 (0.0006)	1.5011 (0.0001)	1.4999 (0.0000)	1.4250 (0.0057)
	1.5	1.5011 (0.0004)	1.5001 (0.0003)	1.4999 (0.0000)	1.3351 (0.0274)
	3	1.4998 (0.0009)	1.4996 (0.0008)	1.5002 (0.0000)	1.2767 (0.0501)
10000	0.5	1.5017 (0.0003)	1.4997 (0.0000)	1.4996 (0.0000)	1.4236 (0.0059)
	1.5	1.5025 (0.0003)	1.5027 (0.0002)	1.5001 (0.0000)	1.3375 (0.0265)
	3	1.5016 (0.0004)	1.5021 (0.0004)	1.5002 (0.0000)	1.2778 (0.0495)

TABLE 6. Estimation results for Cauchy, Gaussian error. Mean estimated values of the four estimators $\widehat{\theta}_{N_c}$, $\widehat{\theta}_{SC_c}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$ are presented for various values of n (1000, 5000 or 10000) and s2n (0.5, 1.5, 3). True value is $\theta^0 = 1.5$. MSE are given in brackets.

function chosen for the construction of $\widehat{\theta}_{N_c}$ depends on σ_ε^2. This estimator is thus more adaptive to changes in $s2n$.

5. A more general estimator

For a large number of regression functions, a weight function w such as the one involved in the definition of the estimator $\hat{\theta}$ can be easily exhibited. Nevertheless for some specific regression functions, it seems not straightforward to find a weight function such that $(wf_\theta)^*/f_\varepsilon^*$ and $(wf_\theta^2)^*/f_\varepsilon^*$ are integrable. We refer to Butucea and Taupin (2008) for a more complete discussion on this subject. Therefore, we propose a generalization of this estimator to relax these conditions.

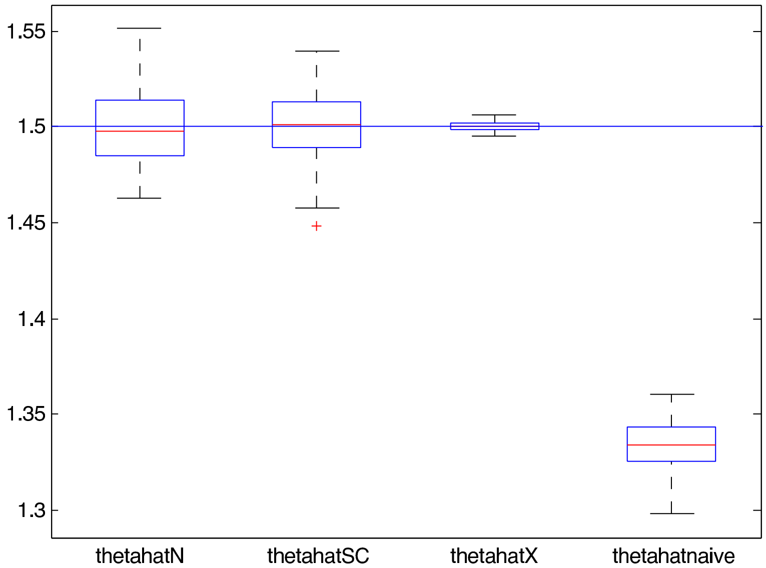

FIGURE 3. Results for Cauchy and Gaussian error, with $n = 5000$ and $\sigma_\varepsilon^2/\mathrm{Var}(X) = 1.5$. Box plots of the four estimators $\widehat{\theta}_{N_c}$, $\widehat{\theta}_{SC_c}$, $\widehat{\theta}_X$ and $\widehat{\theta}_{naive}$, from left to right, based on 100 replications. True value is 1.5 (horizontal line).

5.1. **Definition of the general estimator.** The key idea for this construction is the following. We introduce a density deconvolution kernel K_{n,C_n} defined via its Fourier transform K^*_{n,C_n} by

$$K^*_{n,C_n}(t) = \frac{K^*(t/C_n)}{f^*_\varepsilon(-t)} := \frac{K^*_{C_n}(t)}{f^*_\varepsilon(-t)}, \tag{5.20}$$

where K^* is the Fourier transform of a kernel K and C_n is a sequence which tends to infinity with n. The kernel K belongs to $\mathbb{L}^2(\mathbb{R})$. Its Fourier transform K^* is compactly supported and satisfies $|1 - K^*(t)| \leq \mathbb{1}_{|t|\geq 1}$. Then, for any integrable function Φ, one has $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \Phi \star K_{n,C_n}(Z_i) = \mathbb{E}(\Phi(X))$. Hence we estimate $\mathbb{E}(\Phi(X))$ by $n^{-1}\sum_{i=1}^n \Phi \star K_{n,C_n}(Z_i)$ instead of $n^{-1}\sum_{i=1}^n \Phi(X_i)$ which is not available. We then propose to estimate $S_{\theta^0,P_X}(\theta)$ by

$$\begin{aligned} S_n(\theta) &= \frac{1}{n}\sum_{i=1}^n \mathbb{R}e\left[\left((Z_i - f_\theta)^2 w\right) \star K_{n,C_n}(Z_{i-1})\right] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{R}e\int (Z_i - f_\theta(x))^2\, w(x)\, K_{n,C_n}(Z_{i-1} - x)dx. \end{aligned} \tag{5.21}$$

Using this more general empirical criterion we propose to estimate θ^0 by

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} S_n(\theta). \tag{5.22}$$

Note that the general construction relies to a truncation of integrals in (2.4). Also note that this general construction still works under Conditions $(\mathbf{C_1})$-$(\mathbf{C_5})$. It suffices to chose $K^*(t/C_n) = \mathbb{1}_{|t|\leq C_n}$ with $C_n = +\infty$.

5.2. **Asymptotic properties under general assumptions.** This section presents the asymptotic properties of $\widehat{\theta}$ defined by (5.22) under milder conditions than conditions $(\mathbf{C_1})$-$(\mathbf{C_5})$, when one cannot exhibit a weight function w ensuring that these conditions hold. In this context the estimator is still consistent, but with a rate which is not necessarily the parametric rate. For the sake of simplicity we only consider the case of α-mixing Markov chains.

We assume that

$(\mathbf{A_3})$ On Θ^0, the quantity $w^2(X_0)(Z_1 - f_\theta(X_0))^4$ and the absolute values of its derivatives with respect to θ up to order 2 have a finite expectation.

$(\mathbf{A_4})$ The quantity $\sup_n \sup_{j\in\{1,\cdots,d\}} \mathbb{E}\Big(\sup_{\theta\in\Theta^\circ}\Big|\frac{\partial}{\partial\theta_j}S_n(\theta)\Big|\Big)$ is finite.

$(\mathbf{A_5})$ $\sup_{\theta\in\Theta}|wf_\theta|$, $|w|$ and $\sup_{\theta\in\Theta}|wf_\theta^2|$ belong to $\mathbb{L}_1(\mathbb{R})$.

We say that a function $\psi \in \mathbb{L}_1(\mathbb{R})$ satisfies (5.23) if for a sequence C_n we have

$$\min_{q=1,2} \parallel \psi^*(K^*_{C_n} - 1) \parallel_q^2 + n^{-1} \min_{q=1,2} \left\|\frac{\psi^* K^*_{C_n}}{f^*_\varepsilon}\right\|_q^2 = o(1). \tag{5.23}$$

Theorem 5.1. *Under the assumptions $(\mathbf{I1_1})$, $(\mathbf{I1_2})$, $(\mathbf{N_1})$, $(\mathbf{A_1})$ $(\mathbf{A_3})$ - $(\mathbf{A_5})$, let $\widehat{\theta}$ be defined by (5.22) with C_n such that (5.23) holds for w, wf_θ and wf_θ^2 and their first derivatives with respect to θ. Assume that the sequence (X_k) is α-mixing that is*

$$\alpha_{\mathbf{X}}(k) \underset{n\to\infty}{\longrightarrow} 0, \text{ as } k \underset{n\to\infty}{\longrightarrow} \infty.$$

Then $\mathbb{E}(\|\widehat{\theta} - \theta^0\|^2_{\ell^2}) = o(1)$, as $n \to \infty$ and $\widehat{\theta}$ is a consistent estimator of θ^0.

We now give upper bounds for the rates of convergence under two different types of assumptions:

$(\mathbf{A_6})$ X_0 admits a density f_X with respect to the Lebesgue measure and there exist two constants $C_1(f^2_{\theta^0})$ and $C_2(f_{\theta^0})$ such that $\parallel f_{\theta^0} f_X \parallel_2^2 \leq C_1(f_{\theta^0})$, and $\parallel f^2_{\theta^0} f_X \parallel_2^2 \leq C_2(f^2_{\theta^0})$.

$(\mathbf{A_7})$ $\sup_{z\in\mathbb{R}}\mathbb{E}[f^2_{\theta^0}(X_0)f_\varepsilon(z - X_0)]$ and $\sup_{z\in\mathbb{R}}\mathbb{E}[f_\varepsilon(z - X_0)]$ are finite.

These two assumptions are mostly required for technical reasons. The following theorem still holds when X_0 does not admit a density, under a slightly different moment assumption.

Theorem 5.2. *Suppose that the assumptions of Theorem 5.1 hold. Assume moreover that the sequence $(X_k)_{k\geq 0}$ is α-mixing with $\sum_{k\geq 1}\sqrt{\alpha_{\mathbf{X}}(k)} < \infty$, and that, for all $\theta \in \Theta$, the functions w, $f_\theta w$ and $f_\theta^2 w$ and their derivatives up to order 3 with respect to θ satisfy (5.23).*

1) *Assume that the sequence X_0 admits a density with respect to the Lebesgue measure and that Assumption $(\mathbf{A_6})$ holds. Then $\widehat{\theta} - \theta^0 = O_p(\varphi_n^2)$ with $\varphi_n = \|(\varphi_{n,j})\|_{\ell^2}$, $\varphi^2_{n,j} = B^2_{n,j} + V_{n,j}/n$, $j = 1\ldots, d$, where*

$$B_{n,j} = \min\left\{B^{[1]}_{n,j}, B^{[2]}_{n,j}\right\} \text{ and } V_{n,j}{=}\min\left\{V^{[1]}_{n,j}, V^{[2]}_{n,j}\right\}$$

and for $q = 1, 2$

$$B^{[q]}_{n,j} \;=\; \Big\|(wf^{(1)}_{\theta,j})^*(K^*_{C_n}-1)\Big\|^2_q + \Big\|(wf_{\theta^0}f^{(1)}_{\theta^0,j})^*(K^*_{C_n}-1)\Big\|^2_q,$$

and

$$V^{[q]}_{n,j} \;=\; \Big\|(wf^{(1)}_{\theta^0,j})^*\frac{K^*_{C_n}}{f^*_\varepsilon}\Big\|^2_q + \Big\|(wf_{\theta^0}f^{(1)}_{\theta^0,j})^*\frac{K^*_{C_n}}{f^*_\varepsilon}\Big\|^2_q.$$

2) Assume that ($\mathbf{A_7}$) *holds. Then* $\widehat{\theta}-\theta^0 = O_p(\varphi_n^2)$ *with* $\varphi_n = \|(\varphi_{n,j})\|_{\ell^2}$, $\varphi^2_{n,j} = B^2_{n,j}+V_{n,j}/n$, $j = 1\ldots, d$, *where* $B_{n,j} = B^{[1]}_{n,j}$ *and* $V_{n,j} = \min\left\{V^{[1]}_{n,j}, V^{[2]}_{n,j}\right\}$.

This theorem states an upper bound for the quadratic risk under very general conditions. It holds under mild conditions on w, f_θ and f_ε. We refer to Table 1 in Butucea and Taupin (2008) for more details on the resulting rates.

Appendix A. Properties of the dependence coefficients and examples

A.1. **Covariance inequalities and coupling.** The following results are the key arguments to prove the asymptotic normality of $\widehat{\theta}$. We keep the same notations as in Definition 3.1.

We first recall a covariance inequality due to Rio (1993). For any positive random variable Z, let Q_Z be the inverse cadlag of the tail function $t \to \mathbb{P}(Z > t)$. Let X and Y be two real valued random variables such that $\mathrm{Cov}(X, Y)$ is well defined. The following inequality holds

$$|\mathrm{Cov}(Y, X)| \le 4\int_0^{\alpha(\sigma(Y),\sigma(X))} Q_{|X|}(u)Q_{|Y|}(u)du\,. \tag{A.24}$$

Next, we recall the coupling properties of τ (see Dedecker and Prieur (2005)): enlarging Ω if necessary, there exists X^* distributed as X and independent of $\mathcal{M}$ such that

$$\tau(\mathcal{M}, X) = \mathbb{E}(\|X - X^*\|_{\mathbb{B}})\,. \tag{A.25}$$

A.2. **Dependence properties of autoregressive models.** We recall here the mixing properties of the autoregressive models

$$X_i = f_{\theta^0}(X_{i-1}) + \xi_i,$$

that have been described in particular in the papers by Mokkadem (1985) and Ango-Nzé (1998). For instance, assume that

- the law of ξ_0 has a density f_ξ such that $f_\xi > c > 0$ on a neighborhood of zero, and there exists $S \ge 1$ such that $\mathbb{E}(|\xi_0|^S) < \infty$.
- f_{θ^0} is continuous and there exist $R \ge 1$ and $\rho \in]0, 1[$ such that: for any $|x| \ge R$, $|f_{\theta^0}(x)| \le \rho|x|$.

Then there exists a unique invariant probability measure, and the stationary Markov chain $(X_i)_{i\ge0}$ satisfies $\alpha_{\mathbf{X}}(k) = O(\kappa^k)$ for any $\kappa \in]\rho, 1[$ and is α-mixing.
Now if the second point is weakened to

- f_{θ^0} is continuous and there exist $R \ge 1$ and $\delta \in]0, 1[$ such that: for any $|x| \ge R$, $|f_{\theta^0}(x)| \le |x|(1 - |x|^{-\delta})$.

Then there exists a unique invariant probability measure, and the stationary Markov chain $(X_i)_{i\geq 0}$ satisfies $\alpha_{\mathbf{X}}(k) = O(k^{1-S/\delta})$ and is α-mixing.

Now, if we do not assume that ξ_0 has a density, then the chain may not be α-mixing (and not even irreducible). However, under appropriate assumptions on f_{θ^0}, it is still possible to obtain upper bounds for the coefficient τ. For instance assume that

- there exists $S \geq 1$ such that $\mathbb{E}(|\xi_0|^S) < \infty$.
- $|f_{\theta^0}(x) - f_{\theta^0}(y)| \leq \rho|x-y|$ for some $\rho \in]0,1[$.

Then there exists a unique invariant probability measure, and the stationary Markov chain $(X_i)_{i\geq 0}$ satisfies $\tau_{\mathbf{X},2}(k) = O(\rho^k)$ and is τ-dependent. Now if the second point is weakened to

- there exist δ in $[0,1[$ and C in $]0,1]$ such that $|f'(t)| \leq 1 - C(1+|t|)^{-\delta}$ almost everywhere.

Then there exists a unique invariant probability measure, and for $S > 1+\delta$ the stationary Markov chain $(X_i)_{i\geq 0}$ satisfies $\tau_{\mathbf{X},2}(n) = O(n^{(\delta+1-S)/\delta})$ and is τ-dependent.

Appendix B. proofs of Theorems

B.1. **Proof of Theorem 3.1.** The main point of the proof consists in showing the two following points

i) for any θ in Θ, $S_n(\theta) \underset{n\to\infty}{\xrightarrow{\mathbb{L}^1}} S_{\theta^0,P_X}(\theta)$, with $S_{\theta^0,P_X}(\theta)$ admitting a unique minimum in $\theta = \theta^0$.

ii) For $\omega_2(n,\rho)$ defined as $\omega_2(n,\rho) = \sup\{|S_n(\theta) - S_n(\theta')| : \|\theta-\theta'\|_{\ell^2} \leq \rho\}$, there exists a sequence ρ_k tending to 0, such that

$$\mathbb{E}(\omega_2(n,\rho_k)) = O(\rho_k). \tag{B.1}$$

Let us start with the proof of i) by writing that

$$S_n(\theta) = \frac{1}{n}\sum_{k=1}^{n}\Psi(Z_k, Z_{k-1}), \text{ with } \Psi(Z_1,Z_0) = \frac{1}{2\pi}\mathbb{R}e\int \frac{\left(\left(Z_1 - f_\theta\right)^2 w\right)^*(t)e^{-itZ_0}}{f_\varepsilon^*(-t)}dt,$$

that is seen as a function of a strictly stationary and ergodic sequence of random variables. By the ergodic theorem and Assumption $(\mathbf{A_2})$ we conclude that for any $\theta \in \Theta$,

$$S_n(\theta) \underset{n\to\infty}{\xrightarrow{\mathbb{L}^1}} \mathbb{E}(\psi(Z_1,Z_0)) = S_{\theta^0,P_X}(\theta).$$

It remains now to check that there exists a sequence ρ_k tending to 0, such that (B.1) holds. This follows by the assumption $(\mathbf{C_2})$ and by writing that

$$\sup_{\|\theta-\theta'\|_{\ell^2}\leq\rho} |S_n(\theta) - S_n(\theta')| \leq \sup_{\|\theta-\theta'\|_{\ell^2}\leq\rho} \|\ \theta-\theta'\ \|_{\ell^2} \sup_{\theta\in\Theta^0} \|\ S_n^{(1)}(\theta)\ \|_{\ell^2}\ . \tag{B.2}$$

$\square$

B.2. **Proof of Theorem 3.2.** By using a Taylor expansion based on the smoothness properties of $\theta \mapsto wf_\theta$ and the consistency of $\widehat{\theta}$, we obtain

$$0 = S_n^{(1)}(\widehat{\theta}) = S_n^{(1)}(\theta^0) + S_n^{(2)}(\theta^0)(\widehat{\theta} - \theta^0) + R_n(\widehat{\theta} - \theta^0),$$

with R_n defined by

$$R_n = \int_0^1 [S_n^{(2)}(\theta^0 + s(\widehat{\theta} - \theta^0)) - S_n^{(2)}(\theta^0)]ds. \tag{B.3}$$

This implies that

$$\widehat{\theta} - \theta^0 = -[S_n^{(2)}(\theta^0) + R_n]^{-1} S_n^{(1)}(\theta^0). \tag{B.4}$$

Consequently, we have to check the three following points.

i) $\sqrt{n} S_n^{(1)}(\theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_{0,1})$;

ii) $S_n^{(2)}(\theta^0) \xrightarrow[n\to\infty]{\mathbb{P}} S_{\theta^0,P_X}^{(2)}(\theta^0)$;

iii) R_n defined in (B.3) satisfies $R_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Note that the covariance matrix $\Sigma_{0,1}$ in i) satisfies $\Sigma_{0,1} = \Sigma/4\pi^2$, with Σ defined by the equation (B.6) below. Consequently, according to ii) and iii), the covariance matrix Σ_1 satisfies

$$\Sigma_1 = \frac{1}{4\pi^2}(S_{\theta^0,P_X}^{(2)}(\theta^0))^{-1}\Sigma(S_{\theta^0,P_X}^{(2)}(\theta^0))^{-1}, \quad \text{with } \Sigma \text{ defined by (B.6)}. \tag{B.5}$$

Proof of i)

Under Assumption ($\mathbf{C_2}$),

$$\left(\sqrt{n}S_n^{(1)}(\theta^0)\right)_i = \frac{1}{2\pi\sqrt{n}}\sum_{k=1}^n \mathbb{R}e \int \left(\frac{\partial}{\partial\theta_i}((Z_k - f_\theta)^2)w\Big|_{\theta=\theta^0}\right)^*(t)\frac{e^{-itZk-1}}{f_\varepsilon^*(-t)}dt.$$

We have thus to prove that

$$\frac{1}{2\pi\sqrt{n}}\sum_{k=1}^n \mathbb{R}e\int \big(-2(Z_k - f_{\theta^0})f_{\theta^0}^{(1)}w\big)^*(t)\frac{e^{-itZk-1}}{f_\varepsilon^*(-t)}dt \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma_{0,1}).$$

We first use that $\mathbb{E}(S_n(\theta)) = S_{\theta^0,P_X}(\theta)$ and thus $\mathbb{E}(S_n^{(1)}(\theta^0)) = S_{\theta^0,P_X}^{(1)}(\theta^0) = 0$. Next we write

$$\sqrt{n}S_n^{(1)}(\theta^0) = \sqrt{n}S_n^{(1)}(\theta^0) - \mathbb{E}[\sqrt{n}S_n^{(1)}(\theta^0)] = \frac{1}{2\pi\sqrt{n}}\sum_{k=1}^n T_k$$

with $T_k = -2W_{k,1} + 2W_{k,2}$, and

$$\begin{aligned} W_{k,1} &= Z_k\mathbb{R}e\int \big(f_{\theta^0}^{(1)}w\big)^*(t)\frac{e^{-itZ_{k-1}}}{f_\varepsilon^*(-t)}dt - \mathbb{E}\left[Z_k\mathbb{R}e\int \big(f_{\theta^0}^{(1)}w\big)^*(t)\frac{e^{-itZ_{k-1}}}{f_\varepsilon^*(-t)}dt\right] \\ W_{k,2} &= \mathbb{R}e\int \big(f_{\theta^0}f_{\theta^0}^{(1)}w\big)^*(t)\frac{e^{-itZ_{k-1}}}{f_\varepsilon^*(-t)}dt - \mathbb{E}\left[\mathbb{R}e\int \big(f_{\theta^0}f_{\theta^0}^{(1)}w\big)^*(t)\frac{e^{-itZ_{k-1}}}{f_\varepsilon^*(-t)}dt\right]. \end{aligned}$$

Let $\mathcal{M}_1 = \sigma(X_0, X_1, \varepsilon_0, \varepsilon_1)$. According to Dedecker and Rio (2000), $n^{-1/2}\sum_{k=1}^n T_k$ converges to a centered Gaussian vector with covariance matrix

$$\Sigma = \mathrm{Cov}(T_1, T_1) + 2\sum_{k>1}\mathrm{Cov}(T_1, T_k)\,, \tag{B.6}$$

as soon as for any (p,q) in $\{1,\cdots,d\}\times\{1,\cdots,d\}$

$$\sum_{k=3}^{\infty}\mathbb{E}|(T_1)_p\mathbb{E}((T_k)_q|\mathcal{M}_1)| < \infty. \tag{B.7}$$

For any (p,q) in $\{1,\cdots,d\}\times\{1,\cdots,d\}$ and any $i,j\in\{1,2\}$, we shall give an upper bound for

$$\mathbb{E}\,|(W_{1,i})_p\mathbb{E}((W_{k,j})_q|\mathcal{M}_1)|\,.$$

We first notice that the sequence $(\varepsilon_k, \varepsilon_{k-1})$ is independent of $\mathcal{M}_1\vee\sigma(X_k, X_{k-1})$. It follows that for $i,j\in\{1,2\}$,

$$\mathbb{E}\,|(W_{1,i})_p\mathbb{E}((W_{k,j})_q|\mathcal{M}_1)| = \mathbb{E}\left|(W_{1,i})_p\mathbb{E}((\tilde{W}_{k,j})_q|\mathcal{M}_1)\right|,$$

with

$$\begin{aligned}
(\tilde{W}_{k,1})_q &= X_k\int\left(f^{(1)}_{\theta^0,q}w\right)^*(t)e^{-itX_{k-1}}dt - \mathbb{E}\left[X_k\int\left(f^{(1)}_{\theta^0,q}w\right)^*(t)e^{-itX_{k-1}}dt\right]\\
(\tilde{W}_{k,2})_q &= \int\left(f_{\theta^0}f^{(1)}_{\theta^0,q}w\right)^*(t)e^{-itX_{k-1}}dt - \mathbb{E}\left[\int\left(f_{\theta^0}f^{(1)}_{\theta^0,q}w\right)^*(t)e^{-itX_{k-1}}dt\right].
\end{aligned}$$

Next, since $\mathbb{P}_{(X_{k-1},X_k)|\sigma(\varepsilon_0,\varepsilon_1,X_0,X_1)} = \mathbb{P}_{(X_{k-1},X_k)|\sigma(X_1)}$, we infer that

$$\mathbb{E}\,|(W_{1,i})_p\mathbb{E}((W_{k,j})_q|\mathcal{M}_1)| = \mathbb{E}\left|(W_{1,i})_p\mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right|.$$

Next we use that under Condition $(\mathbf{C_2})$,

$$\begin{aligned}
|(W_{1,1})_p| &\le |Z_1|\int\left|\left(f^{(1)}_{\theta^0,p}w\right)^*(t)\frac{e^{-itZ_0}}{f^*_\varepsilon(-t)}\right|dt + \mathbb{E}\left\{|Z_1|\int\left|\left(f^{(1)}_{\theta^0,p}w\right)^*(t)\frac{e^{-itZ_0}}{f^*_\varepsilon(-t)}\right|dt\right\}\\
&\le |Z_1|\int\left|\left(f^{(1)}_{\theta^0,p}w\right)^*(t)\frac{1}{f^*_\varepsilon(-t)}\right|dt + \mathbb{E}\left\{|Z_1|\int\left|\left(f^{(1)}_{\theta^0,p}w\right)^*(t)\frac{1}{f^*_\varepsilon(-t)}\right|dt\right\}\\
&\le C_1(|Z_1| + \mathbb{E}(|Z_1|)).
\end{aligned}$$

In the same way we get that $|(W_{1,2})_p| \le C_2$.

Now, since ε_1 is independent of X_1, for $j\in\{1,2\}$

$$\begin{aligned}
\mathbb{E}\left|(W_{1,1})_p\mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right| &\le C_1\mathbb{E}\left[(|Z_1| + \mathbb{E}(|Z_1|))\left|\mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right|\right]\\
&\le C\mathbb{E}\left[(|X_1| + \mathbb{E}(|X_1|))\left|\mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right|\right]. \qquad \text{(B.8)}
\end{aligned}$$

In the same way

$$\mathbb{E}\left|(W_{1,2})_p\mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right| \le C\mathbb{E}\left|\mathbb{E}((\tilde{W}_{k,j})_q|X_1)\right|. \tag{B.9}$$

Note that

$$\mathbb{E}\left[(|X_1| + \mathbb{E}(|X_1|))\left|\mathbb{E}((\tilde{W}_{k,1})_q|X_1)\right|\right] = \mathrm{Cov}((|X_1| + \mathbb{E}(|X_1|))\mathrm{sign}(\mathbb{E}((\tilde{W}_{k,1})_q|X_1)), (\tilde{W}_{k,1})_q).$$

Now, we use the covariance inequality (A.24). Note first that

$$(|X_1| + \mathbb{E}(|X_1|))\mathrm{sign}(\mathbb{E}((\tilde{W}_{k,1})_q|X_1)) \leq |X_1| + \mathbb{E}(|X_1|)$$

and

$$|(\tilde{W}_{1,1})_q| \leq D(|X_1| + \mathbb{E}(|X_1|))\,.$$

Since $(X_i)_{i\geq 0}$ is a strictly stationary Markov chain, it is well known that

$$\alpha(\sigma(X_1), \sigma(X_{k-1}, X_k)) = \alpha(\sigma(X_1), \sigma(X_{k-1})) = \alpha_{\mathbf{X}}(k-2)\,. \tag{B.10}$$

Hence, applying (A.24),

$$\mathbb{E}\left|(W_{1,1})_p \mathbb{E}((\tilde{W}_{k,1})_q|X_1)\right| \leq C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^2_{|X_1|}(u)du\,.$$

We conclude that

$$\sum_{k\geq 3}\mathbb{E}\,|(W_{1,1})_p\mathbb{E}((W_{k,1})_q|\mathcal{M}_1)| \leq C\sum_{k\geq 3}\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^2_{|X_1|}(u)du.$$

Finally, using similar arguments for the three quantities $\sum_{k\geq 3}\mathbb{E}\,|(W_{1,2})_p\mathbb{E}((W_{k,1})_q|\mathcal{M}_1)|$, $\sum_{k\geq 3}\mathbb{E}\,|(W_{1,1})_p\mathbb{E}((W_{k,2})_q|\mathcal{M}_1)|$ and $\sum_{k\geq 3}\mathbb{E}\,|(W_{1,2})_p\mathbb{E}((W_{k,2})_q|\mathcal{M}_1)|$ we conclude that

$$\sqrt{n}S_n^{(1)}(\theta^0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \Sigma/(4\pi^2))$$

as soon as

$$\sum_{k\geq 1}\int_0^{\alpha_{\mathbf{X}}(k)} Q^2_{|X_1|}(u)du < \infty.$$

□

Proof of ii)

Under Condition ($\mathbf{C_3}$), for $j, k = 1, \cdots, d$,

$$\left(S_n^{(2)}(\theta)\right)_{j,k} = \frac{1}{2\pi n}\sum_{\ell=1}^n \mathbb{R}e\int\left(-2Z_\ell\frac{\partial^2}{\partial\theta_j\partial\theta_k}(f_\theta w) + \frac{\partial^2}{\partial\theta_j\partial\theta_k}(f_\theta^2 w)\right)^*(t)\frac{e^{-itZ_{\ell-1}}}{f_\varepsilon^*(-t)}dt \tag{B.11}$$

and by applying the ergodic theorem we get that

$$S_n^{(2)}(\theta^0) \xrightarrow[n\to\infty]{\mathbb{P}} S^{(2)}_{\theta^0,P_X}(\theta^0).$$

□

Proof of iii)

Starting from (B.3) and (B.11), the point iii) follows from the assumption ($\mathbf{C_4}$) on the properties of the derivatives at order 3 of wf_θ and wf_θ^2. □

B.3. **Proof of Theorem 3.3.** We follow the proof of Theorem 3.2 and keep the same notations. We have to check that the condition (B.7) holds. We start from the inequalities (B.8) and (B.9). For clarity, let us write

$$(\tilde{W}_{k,1})_q = (\tilde{W}_{k,1})_q(X_k, X_{k-1})\,.$$

Let ψ_M be the truncating function defined by $\psi_M(x) = (x \wedge M) \vee (-M)$. Applying (A.25), let (X_k^*, X_{k-1}^*) be the random variable distributed as (X_k, X_{k-1}) and independent of X_1 such that

$$\frac{1}{2}(\|X_k - X_k^*\|_1 + \|X_{k-1} - X_{k-1}^*\|_1) = \tau(\sigma(X_1), (X_{k-1}, X_k)) \le \tau_{X,2}(k-2)\,.$$

Define the constants K_1 and K_2 by

$$K_1 = \int \left|\left(f^{(1)}_{\theta^0,q} w\right)^*(t)\right| dt < \infty\,, \quad K_2 = \int |t| \left|\left(f^{(1)}_{\theta^0,q} w\right)^*(t)\right| dt < \infty\,.$$

Clearly

$$|X_1 \mathbb{E}((\tilde{W}_{k,1})_q(X_k, X_{k-1})|X_1)| \le M|\mathbb{E}((\tilde{W}_{k,1})_q(X_k, X_{k-1})|X_1)| + K_2|X_1|\mathbf{1}_{|X_1|>M}(|X_k| + \mathbb{E}(|X_k|))\,.$$

Now, since (X_k^*, X_{k-1}^*) is independent of X_1, one has that

$$|\mathbb{E}((\tilde{W}_{k,1})_q(X_k, X_{k-1})|X_1)| = |\mathbb{E}((\tilde{W}_{k,1})_q(X_k, X_{k-1}) - (\tilde{W}_{k,1})_q(X_k^*, X_{k-1}^*)|X_1)|\,.$$

By definition of $(\tilde{W}_{k,1})_q(X_k, X_{k-1})$, there exists a constant C such that

$$\begin{aligned} |(\tilde{W}_{k,1})_q(X_k, X_{k-1}) - (\tilde{W}_{k,1})_q(X_k^*, X_{k-1}^*) - ((\tilde{W}_{k,1})_q(\psi_M(X_k), X_{k-1}) - (\tilde{W}_{k,1})_q(\psi_M(X_k^*), X_{k-1}^*))| \\ \le C(|X_k|\mathbf{1}_{|X_k|>M} + |X_k^*|\mathbf{1}_{|X_k^*|>M}). \end{aligned}$$

Hence

$$\begin{aligned} |\mathbb{E}((\tilde{W}_{k,1})_q(X_k, X_{k-1})|X_1)| &\le |\mathbb{E}((\tilde{W}_{k,1})_q(\psi_M(X_k), X_{k-1}) - (\tilde{W}_{k,1})_q(\psi_M(X_k^*), X_{k-1}^*)|X_1)| \\ &+ C(|X_k|\mathbf{1}_{|X_k|>M} + |X_k^*|\mathbf{1}_{|X_k^*|>M})\,. \end{aligned}$$

Since ψ_M is 1-Lipschitz and bounded by M, and since $x \to \exp(itx)$ is $|t|$-Lipschitz and bounded by 1, under Condition $(\mathbf{C_5})$, one has

$$|(\tilde{W}_{k,1})_q(\psi_M(X_k), X_{k-1}) - (\tilde{W}_{k,1})_q(\psi_M(X_k^*), X_{k-1}^*)| \le MK_2|X_{k-1} - X_{k-1}^*| + K_1|X_k - X_k^*|\,.$$

It follows that

$$\begin{aligned} |X_1 \mathbb{E}((\tilde{W}_{k,1})_q(X_k, X_{k-1})|X_1)| &\le K_2|X_1|\mathbf{1}_{|X_1|>M}(|X_k| + \mathbb{E}(|X_k|)) \\ &+ M^2K_2|X_{k-1} - X_{k-1}^*| + MK_1|X_k - X_k^*|) \\ &+ CM(|X_k|\mathbf{1}_{|X_k|>M} + |X_k^*|\mathbf{1}_{|X_k^*|>M})\,. \end{aligned}$$

Using that

$$|X_1|\mathbf{1}_{|X_1|>M}|X_k| \le \frac{3}{2}X_1^2\mathbf{1}_{|X_1|>M} + \frac{1}{2}X_k^2\mathbf{1}_{|X_k|>M}\,,$$

we infer from (B.8) with $j = 1$ that there exists a positive constant K such that

$$\mathbb{E}\left[\left|(W_{1,1})_p \mathbb{E}((\tilde{W}_{k,1})_q|X_1)\right|\right] \le K(L(M^2) + M(M+1)\tau_{\mathbf{X},2}(k-2))\,,$$

where $L(t) = \mathbb{E}(X_0^2 \mathbf{1}_{X_0^2>t})$. Let then $G(t) = t^{-1}L(t)$, and let G^{-1} be the inverse cadlag of G. Choose then $M^2 = G^{-1}(\tau_{\mathbf{X},2}(k-2))$. We obtain that

$$\mathbb{E}\left[\left|(W_{1,1})_p \mathbb{E}((\tilde{W}_{k,1})_q|X_1)\right|\right] \leq 2K(2G^{-1}(\tau_{\mathbf{X},2}(k-2))\tau_{\mathbf{X},2}(k-2)+\sqrt{G^{-1}(\tau_{\mathbf{X},2}(k-2))}\tau_{\mathbf{X},2}(k-2))\,.$$

It follows that

$$\sum_{k\geq 3}\mathbb{E}\left[\left|(W_{1,1})_p \mathbb{E}((\tilde{W}_{k,1})_q|X_1)\right|\right] < \infty \quad \text{as soon as} \quad \sum_{k>0} G^{-1}(\tau_{\mathbf{X},2}(k))\tau_{\mathbf{X},2}(k) < \infty\,.$$

Easier control holds for the other terms in (B.8) and (B.9). Consequently (B.7) holds as soon as (3.7) holds, and the proof is complete.

B.4. **Proof of Theorem 5.1.** The proof of the consistency under the assumptions of Theorem 5.1 is quite different from the proof of the consistency under Conditions $(\mathbf{C_1})$-$(\mathbf{C_2})$ in Theorem 3.1. This comes from the fact that $S_n(\theta)$ is now a triangular array of the form

$$S_n(\theta) = \frac{1}{n}\sum_{k=1}^n \Psi_n(Z_k, Z_{k-1}) \text{ with } \Psi_n(Z_1, Z_0) = \frac{1}{2\pi}\mathbb{R}e\int \frac{\left(\left(Z_1 - f_\theta\right)^2 w\right)^*(t)e^{-itZ_0}K^*_{C_n}(t)}{f^*_\varepsilon(-t)}dt.$$

In this context we show that

i) For all θ in Θ, $\mathbb{E}[(S_n(\theta) - S_{\theta^0,P_X}(\theta))^2] = o(1)$ as $n \to \infty$.

ii) The control (B.1) holds.

Note first that ii) follows from the upper bound (B.2) and Assumption $(\mathbf{A_4})$.

For the proof of i) we check that for all $\theta \in \Theta$,

$$\mathbb{E}[S_n(\theta)] - S_{\theta^0,P_X}(\theta) = o(1) \quad \text{and } \mathrm{Var}(S_n(\theta)) = o(1), \quad \text{as } n \to \infty. \tag{B.12}$$

Proof of the first part of $(B.12)$. Since $Z_0 = X_0 + \varepsilon_0$, with ε_0 independent of (Z_1, X_0), it follows that

$$\mathbb{E}[S_n(\theta)] = \mathbb{E}\left[\mathbb{R}e\left((Z_1 - f_\theta)^2\, w\right) \star K_{n,C_n}(Z_0)\right] = \mathbb{E}\left[\left((Z_1 - f_\theta)^2\, w\right) \star K_{C_n}(X_0)\right]\,,$$

hence

$$\begin{aligned}\mathbb{E}[S_n(\theta)] - S_{\theta^0,P_X}(\theta) \quad &= \quad \frac{1}{2\pi}\iint (f^2_{\theta^0}(x) + \sigma^2_\xi + \sigma^2_\varepsilon)e^{-iux}w^*(u)(K^*_{C_n} - 1)(u)du P_X(dx)\\ &\quad -\frac{1}{\pi}\iint f_{\theta^0}(x)e^{-iux}(f_\theta w)^*(u)(K^*_{C_n} - 1)(u)du P_X(dx)\\ &\quad +\frac{1}{2\pi}\iint e^{-iux}(f^2_\theta w)^*(u)(K^*_{C_n} - 1)(u)P_X(dx)du.\end{aligned}$$

Now, arguing as in Butucea and Taupin (2008) we get that $|\mathbb{E}[S_n(\theta)] - S_{\theta^0,P_X}(\theta)|^2 = o(1)$.

Proof of the second part of $(B.12)$. Using that the Z_i's are strictly stationary we get that

$$\begin{aligned}\mathrm{Var}[S_n(\theta)] &= \mathrm{Var}\left[n^{-1}\sum_{k=1}^{n}\mathbb{R}e\Big[((Z_k-f_\theta)^2\,w)\star K_{n,C_n}(Z_{k-1})\Big]\right]\\ &\le \frac{1}{n}\;\mathrm{Var}\;(A_{1,0})+\frac{2}{n}\sum_{i=2}^{n}|\;\mathrm{Cov}\big(A_{1,0},A_{i,i-1}\big)|\\ &\le \frac{3}{n}\;\mathrm{Var}\;(A_{1,0})+\frac{2}{n}\sum_{k=3}^{n}|\;\mathrm{Cov}\big(A_{1,0},A_{k,k-1}\big)|\end{aligned}$$

with

$$A_{k,k-1}=\mathbb{R}e\Big[((Z_k-f_\theta)^2\,w)\star K_{n,C_n}(Z_{k-1})\Big].$$

Arguing as in Butucea and Taupin (2008) we obtain that $\lim_{n\to\infty} n^{-1}$ Var $(A_{1,0})=0$. It remains to study

$$\frac{1}{n}\sum_{k=3}^{n}|\;\mathrm{Cov}\big(A_{1,0},A_{k,k-1}\big)|.$$

Lemma B.1. *Let* Ψ *such that* $\mathbb{E}(|\Psi(Z)|)<\infty$ *and let* Φ *be an integrable function. Let*

$$B_{k,k-1}=\mathbb{R}\left[e\Psi(Z_k)\Phi\star K_{n,C_n}(Z_{k-1})\right].$$

Then for $k\ge 3$

$$\begin{aligned}\mathrm{Cov}(B_{k,k-1},B_{1,0}) &= \mathrm{Cov}[\Psi(Z_k)\Phi\star K_{C_n}(X_{k-1}),\Psi(Z_1)\Phi\star K_{C_n}(X_0)]\\ &= \frac{1}{(2\pi)^2}\iint \Phi^*(t)\Phi^*(s)\mathrm{Cov}\big(\Psi(Z_k)e^{-itX_{k-1}},\Psi(Z_1)e^{-isX_0}\big)K^*_{C_n}(t)K^*_{C_n}(s)dtds.\end{aligned}$$

Proof of Lemma B.1: By stationarity we write

$$\mathrm{Cov}\big(B_{k,k-1},B_{1,0}\big)=\mathbb{E}(B_{k,k-1}B_{1,0})-\mathbb{E}(B_{k,k-1})\mathbb{E}(B_{1,0})=\mathbb{E}(B_{k,k-1}B_{1,0})-\big(\mathbb{E}(B_{1,0})\big)^2.$$

Now, we use that the sequences $(X_k)_{k\in\mathbb{Z}}$ and $(\varepsilon_k)_{k\in\mathbb{Z}}$ are independent. This implies that (Z_1,X_0) is independent of ε_0 and thus

$$\mathbb{E}(B_{1,0})=\frac{1}{2\pi}\mathbb{R}e\int\Phi^*(t)\mathbb{E}[\Psi(Z_1)e^{-itZ_0}]\frac{K^*_{C_n}(t)}{f^*_\varepsilon(-t)}dt=\frac{1}{2\pi}\int\Phi^*(t)\mathbb{E}[\Psi(Z_1)e^{-itX_0}]K^*_{C_n}(t)dt.$$

In the same way, for $k\ge 3$,

$$\begin{aligned}&\mathbb{E}(B_{k,k-1}B_{1,0})\\ &\quad=\frac{1}{(2\pi)^2}\mathbb{E}\iint\Phi^*(s)\Phi^*(t)\Psi(Z_k)\Psi(Z_1)\mathbb{R}e\Big(e^{-itZ_{k-1}}\frac{K^*_{C_n}(t)}{f^*_\varepsilon(-t)}\Big)\mathbb{R}e\Big(e^{-isZ_0}\frac{K^*_{C_n}(s)}{f^*_\varepsilon(-s)}\Big)dtds\\ &\qquad=\frac{1}{(2\pi)^2}\iint\Phi^*(s)\Phi^*(t)\mathbb{E}(\Psi(Z_k)e^{-itX_{k-1}}\Psi(Z_1)e^{-isX_0})K^*_{C_n}(t)K^*_{C_n}(s)dtds\,,\end{aligned}$$

and the lemma is proved. □

It follows from Lemma B.1 that for $k \geq 3$,

$$\mathrm{Cov}\big(A_{k,k-1}, A_{1,0}\big) = \mathrm{Cov}\Big[\big((Z_k - f_\theta)^2\, w\big) \star K_{C_n}(X_{k-1}), \big((Z_1 - f_\theta)^2\, w\big) \star K_{C_n}(X_0)\Big] = \sum_{i=1}^{9} C_i,$$

with

$$\begin{aligned}
C_1 &= \frac{1}{(2\pi)^2} \iint \mathrm{Cov}(e^{-itX_{k-1}}, e^{-isX_0})(wf_\theta^2)^*(t)(wf_\theta^2)^*(s)K^*_{C_n}(s)K^*_{C_n}(t)dtds,\\
C_2 &= \frac{1}{\pi^2} \iint \mathrm{Cov}(X_k e^{-itX_{k-1}}, X_1 e^{-isX_0})(wf_\theta)^*(t)(wf_\theta)^*(s)K^*_{C_n}(t)K^*_{C_n}(s)dtds,\\
C_3 &= \frac{1}{(2\pi)^2} \iint \mathrm{Cov}[(X_k^2 + \varepsilon_k^2)e^{-itX_{k-1}}, (X_1^2 + \varepsilon_1^2)e^{-isX_0}]w^*(t)w^*(s)K^*_{C_n}(t)K^*_{C_n}(s)dtds,\\
C_4 &= \frac{-1}{2\pi^2} \iint \mathrm{Cov}(X_k e^{-itX_{k-1}}, e^{-isX_0})(wf_\theta)^*(t)(wf_\theta^2)^*(s)K^*_{C_n}(s)K^*_{C_n}(t)dtds,\\
C_5 &= \frac{-1}{2\pi^2} \iint \mathrm{Cov}(e^{-itX_{k-1}}, X_1 e^{-isX_0})(wf_\theta)^*(s)(wf_\theta^2)^*(t)K^*_{C_n}(s)K^*_{C_n}(t)dtds,\\
C_6 &= \frac{1}{(2\pi)^2} \iint \mathrm{Cov}[(X_k^2 + \varepsilon_k^2)e^{-itX_{k-1}}, e^{-isX_0}]w^*(t)(wf_\theta^2)^*(s)K^*_{C_n}(s)K^*_{C_n}(t)dtds,\\
C_7 &= \frac{1}{(2\pi)^2} \iint \mathrm{Cov}[e^{-itX_{k-1}}, (X_1^2 + \varepsilon_1^2)e^{-isX_0}]w^*(s)(wf_\theta^2)^*(t)K^*_{C_n}(s)K^*_{C_n}(t)dtds,\\
C_8 &= \frac{-1}{2\pi^2} \iint \mathrm{Cov}[(X_k^2 + \varepsilon_k^2)e^{-itX_{k-1}}, X_1 e^{-isX_0}]w^*(t)(wf_\theta)^*(s)K^*_{C_n}(s)K^*_{C_n}(t)dtds,\\
C_9 &= \frac{-1}{2\pi^2} \iint \mathrm{Cov}[X_k e^{-itX_{k-1}}, (X_1^2 + \varepsilon_1^2)e^{-isX_0}]w^*(s)(wf_\theta)^*(t)K^*_{C_n}(s)K^*_{C_n}(t)dtds
\end{aligned}$$

Easy computations give

$$\begin{aligned}
&\mathrm{Cov}[(X_k^2 + \varepsilon_k^2)e^{-itX_{k-1}}, (X_1^2 + \varepsilon_1^2)e^{-isX_0}] =\\
&\qquad \mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, X_1^2 e^{-isX_0}) + \sigma_\varepsilon^2 \mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, e^{-isX_0})\\
&\qquad\quad + \sigma_\varepsilon^2 \mathrm{Cov}(e^{-itX_{k-1}}, X_1^2 e^{-isX_0}) + \sigma_\varepsilon^4 \mathrm{Cov}(e^{-itX_{k-1}}, e^{-isX_0}),
\end{aligned}$$

$$\mathrm{Cov}[(X_k^2 + \varepsilon_k^2)e^{-itX_{k-1}}, e^{-isX_0}] = \mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, e^{-isX_0}) + \sigma_\varepsilon^2 \mathrm{Cov}(e^{-itX_{k-1}}, e^{-isX_0}),$$

$$\mathrm{Cov}[(X_k^2 + \varepsilon_k^2)e^{-itX_{k-1}}, X_1 e^{-isX_0}] = \mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, X_1 e^{-isX_0}) + \sigma_\varepsilon^2 \mathrm{Cov}(e^{-itX_{k-1}}, X_1 e^{-isX_0}).$$

which induces the decomposition $\mathrm{Cov}\big(A_{k,k-1}, A_{1,0}\big) = \sum_{i=1}^{9} E_i$, with

$$\begin{aligned}
E_1 = \frac{1}{(2\pi)^2} &\iint \mathrm{Cov}(e^{-itX_{k-1}}, e^{-isX_0})K^*_{C_n}(t)K^*_{C_n}(s)\\
&\times [(wf_\theta^2)^*(t)(wf_\theta^2)^*(s) + \sigma_\varepsilon^4 w^*(t)w^*(s) + \sigma_\varepsilon^2 w^*(t)(wf_\theta)^*(s) + \sigma_\varepsilon^2 w^*(s)(wf_\theta)^*(t))]dtds,
\end{aligned}$$

$$
\begin{aligned}
E_2 &= C_2 = \frac{1}{\pi^2}\iint \mathrm{Cov}(X_k e^{-itX_{k-1}}, X_1 e^{isX_0})(wf_\theta)^*(t)(wf_\theta)^*(s)K^*_{C_n}(t)K^*_{C_n}(s)dtds,\\
E_3 &= \frac{1}{(2\pi)^2}\iint \mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, X_1^2 e^{-isX_0})w^*(t)w^*(s)K^*_{C_n}(t)K^*_{C_n}(s)dtds,\\
E_4 &= \frac{-1}{2\pi^2}\iint \mathrm{Cov}(X_k e^{-itX_{k-1}}, e^{-isX_0})K^*_{C_n}(s)K^*_{C_n}(t)(wf_\theta)^*(t)((wf_\theta^2)^*(s)+\sigma_\varepsilon^2 w^*(s))dtds,\\
E_5 &= \frac{-1}{2\pi^2}\iint \mathrm{Cov}(e^{-itX_{k-1}}, X_1 e^{-isX_0})K^*_{C_n}(s)K^*_{C_n}(t)(wf_\theta)^*(s)((wf_\theta^2)^*(t)+\sigma_\varepsilon^2 w^*(t))dtds,\\
E_6 &= \frac{1}{(2\pi)^2}\iint \mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, e^{-isX_0})K^*_{C_n}(t)K^*_{C_n}(s)w^*(t)(\sigma_\varepsilon^2 w^*(s)+(wf_\theta^2)^*(s))dtds,\\
E_7 &= \frac{1}{(2\pi)^2}\iint \mathrm{Cov}(e^{-itX_{k-1}}, X_1^2 e^{isX_0})K^*_{C_n}(t)K^*_{C_n}(s)w^*(s)(\sigma_\varepsilon^2 w^*(t)+(wf_\theta^2)^*(t))dtds,\\
E_8 &= \frac{-1}{2\pi^2}\iint \mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, X_1 e^{-isX_0})w^*(t)(wf_\theta)^*(s)K^*_{C_n}(s)K^*_{C_n}(t)dtds,\\
E_9 &= \frac{-1}{2\pi^2}\iint \mathrm{Cov}(X_k e^{-itX_{k-1}}, X_1^2 e^{-isX_0})w^*(s)(wf_\theta)^*(t)K^*_{C_n}(s)K^*_{C_n}(t)dtds\,.
\end{aligned}
$$

Using (A.24) and (B.10), we have the upper bounds

$$
\begin{aligned}
|\mathrm{Cov}(e^{-itX_{k-1}}, e^{-isX_0})| &\le C\alpha_{\mathbf{X}}(k-1)\\
|\mathrm{Cov}(X_k e^{-itX_{k-1}}, X_1 e^{-isX_0})| &\le C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^2_{|X|}(u)du\\
|\mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, X_1^2 e^{-isX_0})| &\le C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^4_{|X|}(t)dt\\
|\mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, e^{-isX_0})| &\le C\int_0^{\alpha_{\mathbf{X}}(k-1)} Q^2_{|X|}(t)dt\\
|\mathrm{Cov}(e^{-itX_{k-1}}, X_1^2 e^{-isX_0})| &\le C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^2_{|X|}(t)dt\\
|\mathrm{Cov}(X_k^2 e^{-itX_{k-1}}, X_1 e^{-isX_0})| &\le C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^3_{|X|}(t)dt\\
|\mathrm{Cov}(X_k e^{-itX_{k-1}}, X_1^2 e^{-isX_0})| &\le C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^3_{|X|}(t)dt\,.
\end{aligned}
$$

Since $\mathbb{E}(X_1^4)<\infty$ and $\lim_{k\to\infty}\alpha_{\mathbf{X}}(k)=0$, we infer that $\lim_{k\to\infty}|\operatorname{Cov}\big(A_{k,k-1},A_{1,0}\big)|=0$. Now, by Cesaro's mean convergence theorem

$$
\lim_{n\to\infty}\frac{1}{n}\sum_{k=3}^{n}|\operatorname{Cov}\big(A_{1,0},A_{k,k-1}\big)|=0\,.
$$

This completes the proof of the consistency.

B.5. **Proof of Theorem 5.2.**

Proof of 1) in Theorem 5.2. Starting from the decomposition (B.4) we shall check the three following points.

i) $\mathbb{E}\left[(S_n^{(1)}(\theta^0)-S^{(1)}_{\theta^0,P_X}(\theta^0))(S_n^{(1)}(\theta^0)-S^{(1)}_{\theta^0,P_X}(\theta^0))^\top\right]=O[\varphi_n\varphi_n^\top]$

ii) $S_n^{(2)}(\theta^0)\xrightarrow[n\to\infty]{\mathbb{P}}S^{(2)}_{\theta^0,P_X}(\theta^0)$;

iii) R_n defined in (B.3) satisfies $R_n\xrightarrow[n\to\infty]{\mathbb{P}}0$.

The rate of convergence of $\widehat{\theta}$ is thus given by the order of

$$\mathbb{E}\left[(S_n^{(1)}(\theta^0)-S^{(1)}_{\theta^0,P_X}(\theta^0))(S_n^{(1)}(\theta^0)-S^{(1)}_{\theta^0,P_X}(\theta^0))^\top\right].$$

Proof of i)

We first write

$$\begin{aligned}\left(S_n^{(1)}(\theta)\right)_i &= \frac{1}{n}\sum_{k=1}^n\frac{\partial}{\partial\theta_i}\mathbb{R}e\left[((Z_k-f_\theta)^2w)\star K_{n,C_n}(Z_{k-1})-\mathbb{E}[(Z_k-f_\theta(X_{k-1}))^2w(X_{k-1})]\right]\\ &= \frac{1}{n}\sum_{k=1}^n\left(\frac{\partial}{\partial\theta_i}\mathbb{R}e(Z_k-f_\theta)^2w\star K_{n,C_n}(Z_{k-1})-\mathbb{E}\left[\frac{\partial}{\partial\theta_i}(Z_k-f_\theta(X_{k-1}))^2w(X_{k-1})\right]\right).\end{aligned}$$

Study of the bias. As in Butucea and Taupin (2008), we get that

$$\left|\mathbb{E}\left[\left(S_n^{(1)}(\theta^0)\right)_j\right]\right|\le C_1(f_{\theta^0},w,f_\varepsilon)\min\left[B^{[1]}_{n,j}B^{[2]}_{n,j}\right],$$

for $B^{[q]}_{n,j}$, $q=1,2$, defined in Theorem 5.2.

Study of the variance. For the variance term, note first that

$$\mathrm{Var}\left(\left(S_n^{(1)}(\theta^0)\right)_j\right)\le\frac{3}{n}\mathrm{Var}(D_{1,0})+\frac{2}{n}\sum_{k=3}^n|\mathrm{Cov}(D_{1,0},D_{k,k-1})|,$$

with

$$D_{k,k-1}=\mathbb{R}e\Big(\big(-2Z_kf^{(1)}_{\theta^0,j}+2f_\theta f^{(1)}_{\theta^0,j}\big)w\Big)\star K_{n,C_n}(Z_{k-1}).$$

The first part in $\mathrm{Var}\big[(S_n^{(1)}(\theta^0))_j\big]$ is controlled as in Butucea and Taupin (2008) by

$$\frac{1}{n}\mathrm{Var}(D_{1,0})\le\frac{C(\sigma_\xi^2,f_{\theta^0},f^{(1)}_{\theta^0,j},w,f_\varepsilon)}{n}\min\{V^{[1]}_{n,j}(\theta^0),V^{[2]}_{n,j}(\theta^0)\}\tag{B.13}$$

with $V^{[q]}_{n,j}$, $q=1,2$ defined in Theorem 5.2. We now control the term

$$\frac{1}{n}\sum_{k=3}^n|\mathrm{Cov}(D_{1,0},D_{k,k-1})|.$$

Applying again Lemma B.1, we obtain that

$$\mathrm{Cov}(D_{1,0},D_{k,k-1})=F_1+F_2+F_3+F_4$$

with

$$\begin{aligned}
F_1 &= \frac{1}{\pi^2}\mathbb{R}e \iint \mathrm{Cov}(X_k e^{-itX_{k-1}}, X_1 e^{-isX_0})\big(f^{(1)}_{\theta^0,j}w\big)^*(t)\big(f^{(1)}_{\theta^0,j}w\big)^*(s)K^*_{C_n}(t)K^*_{C_n}(s)dtds\\
F_2 &= \frac{1}{\pi^2}\mathbb{R}e \iint \mathrm{Cov}(e^{-itX_{k-1}}, e^{-isX_0})\big(f_{\theta^0}f^{(1)}_{\theta^0,j}w\big)^*(t)\big(f_{\theta^0}f^{(1)}_{\theta^0,j}w\big)^*(s)K^*_{C_n}(t)K^*_{C_n}(s)dtds\\
F_3 &= \frac{-1}{\pi^2}\mathbb{R}e \iint \mathrm{Cov}(X_k e^{-itX_{k-1}}, e^{-isX_0})\big(f^{(1)}_{\theta^0,j}w\big)^*(t)\big(f_{\theta^0}f^{(1)}_{\theta^0,j}w\big)^*(s)K^*_{C_n}(t)K^*_{C_n}(s)dtds\\
F_4 &= \frac{-1}{\pi^2}\mathbb{R}e \iint \mathrm{Cov}(e^{-itX_{k-1}}, X_1 e^{-isX_0})\big(f^{(1)}_{\theta^0,j}w\big)^*(t)\big(f^{(1)}_{\theta^0,j}w\big)^*(s)K^*_{C_n}(t)K^*_{C_n}(s)dtds.
\end{aligned}$$

Using (A.24) and (B.10) we have the upper bounds

$$\begin{aligned}
|\mathrm{Cov}(e^{-itX_{k-1}}, e^{isX_0})| &\leq C\alpha_{\mathbf{X}}(k-1)\\
|\mathrm{Cov}(X_k e^{-itX_{k-1}}, X_1 e^{isX_0})| &\leq C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q^2_{|X|}(u)du\\
|\mathrm{Cov}(X_k e^{itX_{k-1}}, e^{isX_0})| &\leq C\int_0^{\alpha_{\mathbf{X}}(k-1)} Q_{|X|}(u)du\\
|\mathrm{Cov}(e^{itX_{k-1}}, X_1 e^{isX_0})| &\leq C\int_0^{\alpha_{\mathbf{X}}(k-2)} Q_{|X|}(u)du.
\end{aligned}$$

Since $\mathbb{E}(X_1^4) < \infty$, we infer that $Q_{|X|}(u) \leq Cu^{-1/4}$, and consequently all the covariance terms are $O(\sqrt{\alpha_{\mathbf{X}}(k)})$. Finally, if $\sum_{k>0}\sqrt{\alpha_{\mathbf{X}}(k)} < \infty$, then

$$\frac{1}{n}\sum_{k=3}^{n}|\mathrm{Cov}(D_{1,0}, D_{k,k-1})| \leq \frac{C}{n}\,.$$

This, together with (B.13), implies that

$$\mathrm{Var}\left[\big(S_n^{(1)}(\theta^0)\big)_j\right] \leq \frac{C}{n}\min\{V^{[1]}_{n,j}(\theta^0), V^{[2]}_{n,j}(\theta^0)\}\,.$$

Proof of ii)

The proof of **ii)** starts from the expression of the second derivative of the estimation criterion

(B.14)
$$\Big(S_n^{(2)}(\theta)\Big)_{j,k} = \frac{1}{2\pi n}\sum_{\ell=1}^{n}\mathbb{R}e\int\left(-2Z_\ell\frac{\partial^2}{\partial\theta_j\partial\theta_k}(f_\theta w) + \frac{\partial^2}{\partial\theta_j\partial\theta_k}(f_\theta^2 w)\right)^*(t)\frac{K^*_{C_n}(t)e^{-itZ_{\ell-1}}}{f^*_\varepsilon(-t)}dt.$$

Following the same lines as for the consistency we prove that

$$S_n^{(2)}(\theta^0) \underset{n\to\infty}{\overset{\mathbb{P}}{\longrightarrow}} S^{(2)}_{\theta^0,P_X}(\theta^0).$$

$\square$

Proof of iii)

The proof of **iii)** follows from (B.14), from the smoothness properties of wf_θ and from Assumption ($\mathbf{A_4}$).

$\square$

Proof of 2) in Theorem 5.2. The proof of 2) in theorem 5.2 is quite similar to the proof of 1). The main differences appear in the control of the bias and variance of $S_n^{(1)}(\theta^0)$. More precisely, we start from

$$S_n^{(1)}(\theta) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{R}e\left(\frac{\partial}{\partial\theta}(Z_k - f_\theta)^2 w\right) \star K_{n,C_n}(Z_{k-1}) - \mathbb{E}\left[\frac{\partial}{\partial\theta}(Z_k - f_\theta(X_{k-1}))^2 w(X_{k-1})\right].$$

Study of the bias Since $P_{Z,X}(z,z) = P_X(x)f_\varepsilon(z-x)$ we obtain that $\mathbb{E}[S_n^{(1)}(\theta^0)] - S^{(1)}_{\theta^0,P_X}(\theta^0)$ is equal to

$$-2\mathbb{E}\left[f_{\theta^0}(X_0)(f^{(1)}_{\theta^0}w)\star K_{C_n}(X_0) - f_{\theta^0}(X_0)f^{(1)}_{\theta^0}(X_0)w(X_0)\right] + 2\mathbb{E}\left[(f^{(1)}_{\theta^0}f_{\theta^0}w)\star K_{C_n}(X_0) - (f^{(1)}_{\theta^0}f_{\theta^0}w)(X_0)\right],$$

that is $\mathbb{E}[S_n^{(1)}(\theta^0)] - S^{(1)}_{\theta^0,P_X}(\theta^0)$ is equal to

$$-2\mathbb{R}e\iint f_{\theta^0}(x)e^{-iux}(f^{(1)}_{\theta^0}w)^*(u)(K^*_{C_n}(u)-1)P_X(dx)\,du + 2\mathbb{R}e\iint e^{-iux}(f_{\theta^0}f^{(1)}_{\theta^0}w)^*(u)(K^*_{C_n}(u)-1)P_X(dx)\,du.$$

It follows that for $j = 1,\cdots,d$,

$$\left|\mathbb{E}[(S_n^{(1)}(\theta^0))_j] - (S^{(1)}_{\theta^0,P_X}(\theta^0))_j\right| \leq \mathbb{E}|f_{\theta^0}(X_0)|\int|(f^{(1)}_{\theta^0,j}w)^*(u)(K^*_{C_n}(u)-1)|du + \int|(f_{\theta^0}f^{(1)}_{\theta^0,j}w)^*(u)(K^*_{C_n}(u)-1)|du.$$

Study of the variance For the study of the variance we combine the proof in Butucea and Taupin (2008) and the proof of **1)** of Theorem 5.2. For these reasons we only give a sketch of the proof, with details only for specific parts. As for the proof of **1)** we start from

$$\mathrm{Var}\big[(S_n^{(1)}(\theta^0))_j\big] = \frac{1}{n}\mathrm{Var}\left[\mathbb{R}e\left(\frac{\partial[-2Z_kf_\theta w + f_\theta^2 w]}{\partial\theta_j}\Big|_{\theta=\theta^0}\right)\star K_{n,C_n}(Z_{k-1})\right] + \frac{2}{n^2}\sum_{1\leq j<k\leq n}\mathrm{Cov}(D_{k,k-1},D_{j,j-1}),$$

with $D_{k,k-1}$ defined in (B.5). The control of $(2/n^2)\sum_{1\leq j<k\leq n}\mathrm{Cov}(D_{k,k-1},D_{j,j-1})$ is done as in the proof of **1)**. We now control the first part of $\mathrm{Var}\big((S_n^{(1)}(\theta^0))_j\big)$.

$$\mathrm{Var}\big[(S_n^{(1)}(\theta^0))_j\big] \leq \frac{C}{n}\mathbb{R}e\mathbb{E}\left[\left(\frac{\partial[-2Z_if_\theta w + f_\theta^2 w]}{\partial\theta_j}\Big|_{\theta=\theta^0}\right)\star K_{n,C_n}(Z_i)\right]^2.$$

In other words,

$$\begin{aligned}\mathrm{Var}\big[(S_n^{(1)}(\theta^0))_j\big] &\leq \frac{C}{n}\mathbb{R}e\mathbb{E}\left[\left(Z_i f_{\theta^0}^{(1)}w + f_{\theta^0}f_{\theta^0}^{(1)}w\right)\star K_{n,C_n}(Z_i)\right]^2\\ &= \frac{C}{n}\mathbb{R}e\mathbb{E}\left[\left((f_{\theta^0}^2(X_0)+\sigma_\xi^2)f_{\theta^0}^{(1)}w + f_{\theta^0}f_{\theta^0}^{(1)}w\right)\star K_{n,C_n}(Z_0)\right]^2.\end{aligned}$$

Now, write that

$$\mathbb{R}e\mathbb{E}\left[\left((f_{\theta^0}^2(X_0)+\sigma_\xi^2)f_{\theta^0}^{(1)}w + f_{\theta^0}f_{\theta^0}^{(1)}w\right)\star K_{n,C_n}(Z_0)\right]^2 = II_1 + II_2,$$

with

$$\begin{aligned}II_1 &= \mathbb{R}e\iint f_\varepsilon(z-x)(f_{\theta^0}^2(x)+\sigma_\xi^2)\left(\int (f_{\theta^0}^{(1)}w)(u)K_{n,C_n}(z-u)du\right)^2 P_X(dx)dz\\ II_2 &= \mathbb{R}e\iint f_\varepsilon(z-x)\left(\int (f_{\theta^0}f_{\theta^0}^{(1)}w)(u)K_{n,C_n}(z-u)du\right)^2 P_X(dx)dz.\end{aligned}$$

We apply Hölder Inequality and obtain that

$$|II_1| \leq \sup_{z\in\mathbb{R}}\mathbb{E}[(f_{\theta^0}^2(X_0)+\sigma_\xi^2)f_\varepsilon(z-X_0)] \parallel (f_{\theta^0}^{(1)}w)\star K_{n,C_n}\parallel_2^2,$$

and that $|II_1|$ is also less than

$$\mathbb{E}[(f_{\theta^0}^2(X_0)+\sigma_\xi^2)] \parallel (f_{\theta^0}^{(1)}w)\star K_{n,C_n}\parallel_\infty^2$$

In the same way we have

$$|II_2| \leq \sup_{z\in\mathbb{R}}\mathbb{E}[f_\varepsilon(z-X_0)] \parallel (f_{\theta^0}f_{\theta^0}^{(1)}w)\star K_{n,C_n}\parallel_2^2, \text{ and } II_2 \leq \parallel (f_{\theta^0}f_{\theta^0}^{(1)}w)\star K_{n,C_n}\parallel_\infty^2.$$

Consequently we have

$$\text{(B.15)}\quad \mathrm{Var}\big[(S_n^{(1)}(\theta^0))_j\big] \leq \frac{C(\sigma_\xi^2, f_{\theta^0}, f_\varepsilon)}{n}\left[\parallel (f_{\theta^0}^{(1)}w)\star K_{n,C_n}\parallel_2^2 + \parallel (f_{\theta^0}f_{\theta^0}^{(1)}w)\star K_{n,C_n}\parallel_2^2\right]$$

and

$$\text{(B.16)}\quad \mathrm{Var}\big[(S_n^{(1)}(\theta^0))_j\big] \leq \frac{C_1(f_{\theta^0})}{n}\left[\parallel (f_{\theta^0}^{(1)}w)\star K_{n,C_n}\parallel_2^2 + \parallel (f_{\theta^0}f_{\theta^0}^{(1)}w)\star K_{n,C_n}\parallel_1^2\right].$$

By combining (B.15) and (B.16), we get that

$$\mathrm{Var}\big[(S_n^{(1)}(\theta^0))_j\big] \leq \frac{C((f_{\theta^0}, \sigma_\xi^2, f_\varepsilon)}{n}\min\{V_{n,j}^{[1]}(\theta^0), V_{n,j}^{[2]}(\theta^0)\}$$

with $V_{n,j}^{[q]}$, $q=1,2$ defined in Theorem 5.2.

□

References